

From: Sitton, Jehanne Souaya  
Sent: Monday, July 10, 2006 4:38 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search request for 10/723,518

please perform the following sequence search:

please search positions 33601-34000 of SEQ ID NO: 1 where position 33670 is a G, in all nucleic acid databases. please include the first 50 hits and alignments or each.

thanks,  
Jehanne Sitton  
Primary Examiner  
Art Unit 1634  
Rem-2D81 (office)  
Rem-2C70 (mailbox)  
571-272-0752

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2006, 16:57:29 ; Search time 9797 Seconds  
(without alignments)  
2610.898 Million cell updates/sec

Title: SEQ1-33670G

Perfect score: 399.2  
Sequence: 1 aagaagctctctctctata.....taactcagatcagaagcgag 400

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

GenEmbl:\*  
1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_srs:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vl:\*  
11: gb\_ov:\*  
12: gb\_ptg:\*  
13: gb\_in:\*  
14: gb\_om:\*  
15: gb\_ba:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

No.	Score	Query Match	Length	ID	Description
1	397.6	99.6	32745	5	AY675320 Homo sapi
2	397.6	99.6	148386	5	AC104391 Homo sapi
3	397.6	99.6	193100	12	AC016729 Homo sapi
4	397.6	99.6	193363	2	CS200151 Sequence
5	397.6	99.6	193364	5	AC087350 Homo sapi
6	379.6	95.1	156441	5	AP004074 Homo sapi
7	62.2	15.6	211875	12	AC152315 Bos tauru
8	60.8	15.2	256751	6	AC122925 Mus muscu
9	60.6	15.2	120558	11	CR450689 Zebrafish
10	59.6	14.9	152637	12	AC151196 Mus muscu
11	59.6	14.9	181457	6	AC113480 Mus muscu
12	59	14.8	64239	12	AC130372 Homo sapi
13	58.8	14.7	104485	5	AC069439 Homo sapi
14	58.6	14.7	92354	5	AC106720 Homo sapi
15	58.6	14.7	147328	12	AC027755 Homo sapi
16	58.6	14.7	148851	12	AC016227 Homo sapi
17	58.6	14.7	303189	12	AC108728 Homo sapi
18	57.2	14.3	175262	5	AC021491 Homo sapi

19	56.6	14.2	149879	12	AC026572 Homo sapi
20	56.6	14.2	169897	11	AL929213 Zebrafish
21	56.6	14.2	211076	12	AC012214 Homo sapi
22	56.2	14.1	112727	6	AC098683 Mus muscu
23	56.2	14.1	118646	5	AC060796 Homo sapi
24	56.2	14.1	235471	6	AC156029 Mus muscu
25	56.2	14.1	236644	12	AC096353 Rattus no
26	56.2	14.1	239980	12	AC133731 Rattus no
27	56	14.0	319	6	AL585780 Taniarciu
28	56	14.0	171989	5	BS000611 Pan trogl
29	56	14.0	179557	5	AC151479 Pan trogl
30	56	14.0	224014	5	BS000534 Pan trogl
31	56	14.0	232304	5	AC146193 Pan trogl
32	55.8	14.0	60810	5	AC025227 Homo sapi
33	55.8	14.0	110139	11	CR749745 Zebrafish
34	55.8	14.0	212207	5	AC093853 Homo sapi
35	55.6	13.9	176108	5	AC026118 Homo sapi
36	55.6	13.9	184182	5	AC093515 Homo sapi
37	55.2	13.8	668	7	AL593617 Metlagris
38	55.2	13.8	668	7	AL593618 Metlagris
39	55.2	13.8	61255	12	AC093909 Mus muscu
40	55.2	13.8	151358	12	AL590437 Homo sapi
41	55.2	13.8	190080	5	AC022463 Homo sapi
42	55.2	13.8	197131	6	AC136742 Mus muscu
43	55.2	13.8	235887	12	AC106092 Rattus no
44	55	13.8	423	4	AF158547 Pinus tae
45	55	13.8	645	4	AF286606 Pinus tae
46	55	13.8	84422	12	BX088574 Continuation (5 of
47	55	13.8	107344	12	AC127799 Continuation (4 of
48	55	13.8	110000	12	BX088574 Dario rer
49	55	13.8	180358	12	BX323034 Dario rer
50	55	13.8	226227	12	AC133707 Rattus no
51	54.8	13.7	75594	5	AP003093 Homo sapi
52	54.8	13.7	110000	12	AC128221 Rattus no
53	54.8	13.7	122331	5	AL356072 Human DNA
54	54.8	13.7	141529	12	AC025330 Homo sapi
55	54.8	13.7	153605	12	AC019360 Homo sapi
56	54.8	13.7	156481	11	CR376730 Zebrafish
57	54.8	13.7	157708	5	AP005357 Homo sapi
58	54.8	13.7	187098	6	AC166486 Mus muscu
59	54.8	13.7	191692	12	AC023834 Mus muscu
60	54.8	13.7	195959	12	AC034143 Homo sapi
61	54.8	13.7	196490	6	AC124766 Mus muscu
62	54.8	13.7	198254	12	AC016049 Homo sapi
63	54.8	13.7	225253	12	AC098950 Rattus no
64	54.8	13.7	226724	6	AC125199 Mus muscu
65	54.8	13.7	266823	12	AC132688 Rattus no
66	54.8	13.7	271848	12	BX470091 Dario rer
67	54.6	13.7	110000	12	AL928849 Continuation (3 of
68	54.6	13.7	110000	12	AL928849 Continuation (4 of
69	54.6	13.7	110159	6	AC068178 Homo sapi
70	54.6	13.7	187999	6	AL672268 Mouse DNA
71	54.6	13.7	196315	5	AC160873 Colobus g
72	54.6	13.7	211372	12	AC153428 Bos tauru
73	54.6	13.7	223309	6	AC079446 Mus muscu
74	54.6	13.7	241975	11	AL928668 Zebrafish
75	54.4	13.6	124524	5	AC117411 Homo sapi
76	54.4	13.6	149432	12	AC079309 Homo sapi
77	54.4	13.6	175775	5	AC006947 Homo sapi
78	54.4	13.6	175775	5	CR812943 Dario rer
79	54.4	13.6	237236	12	AC111444 Rattus no
80	54.2	13.6	86929	12	AL591666 Dario rer
81	54.2	13.6	162692	5	AC092958 Homo sapi
82	54.2	13.6	166491	12	AC021861 Homo sapi
83	54.2	13.6	221775	12	AC092992 Homo sapi
84	54.2	13.6	221794	12	AC020727 Homo sapi
85	54.2	13.6	228685	12	AC105531 Rattus no
86	54.2	13.6	230223	12	AC125554 Rattus no
87	54.2	13.6	232191	12	AC174022 Bos tauru
88	54.2	13.6	244221	12	AC135828 Rattus no
89	54.2	13.6	266833	12	AC172676 Bos tauru
90	54.2	13.6	301262	12	AC171554 Bos tauru
91	54	13.5	402	7	AB144168 Homo sapi



C	92	54	13.5	78693	5	AC114734	AC114734 Homo sapi
	93	54	13.5	109340	12	CT573349	CT573349 Danio rer
	94	54	13.5	140079	5	AC147286	AC147286 Pan trogl
	95	54	13.5	159649	11	BX927349	BX927349 Zebrafish
	96	54	13.5	161173	12	AC149115	AC149115 Papio anu
	97	54	13.5	165603	12	CT027765	CT027765 Danio rer
	98	54	13.5	172015	12	AC149194	AC149194 Papio anu
C	99	54	13.5	173071	11	CR925789	CR925789 Zebrafish
	100	53.8	13.5	372	7	AB147472	AB147472 Homo sapi
	101	53.8	13.5	45687	5	AC134730	AC134730 Homo sapi
	102	53.8	13.5	54073	12	AC133458	AC133458 Homo sapi
	103	53.8	13.5	114411	5	AC010599	AC010599 Homo sapi
	104	53.8	13.5	152555	5	AL354830	AL354830 Human DNA
	105	53.8	13.5	186524	12	AC024465	AC024465 Homo sapi
	106	53.8	13.5	188587	6	AC124464	AC124464 Mus muscu
C	107	53.8	13.5	200840	5	AL158072	AL158072 Human DNA
	108	53.8	13.5	222879	12	CR931761	CR931761 Danio rer
	109	53.8	13.5	225611	6	AC131724	AC131724 Mus muscu
	110	53.8	13.5	239363	12	AC121057	AC121057 Rattus no
	111	53.6	13.4	134841	5	AC018840	AC018840 Homo sapi
	112	53.6	13.4	145031	5	BS000601	BS000601 Pan trogl
	113	53.6	13.4	161872	12	CT027724	CT027724 Danio rer
	114	53.6	13.4	164376	12	AC010823	AC010823 Homo sapi
C	115	53.6	13.4	168812	5	AC011327	AC011327 Homo sapi
	116	53.6	13.4	170577	5	AC149095	AC149095 Pan trogl
	117	53.6	13.4	172077	5	AC023514	AC023514 Homo sapi
	118	53.6	13.4	173765	5	AC146209	AC146209 Pan trogl
	119	53.6	13.4	173719	11	BX005313	BX005313 Zebrafish
	120	53.6	13.4	180299	5	AC018493	AC018493 Homo sapi
C	121	53.6	13.4	181900	5	AC066606	AC066606 Homo sapi
	122	53.6	13.4	181989	5	BS000663	BS000663 Pan trogl
	123	53.6	13.4	188379	5	AC026200	AC026200 Homo sapi
	124	53.6	13.4	190170	12	AC149721	AC149721 Bos tauru
	125	53.6	13.4	202904	12	AP002077	AP002077 Homo sapi
	126	53.6	13.4	223910	6	AC159712	AC159712 Mus muscu
	127	53.4	13.4	119387	5	AC108128	AC108128 Homo sapi
C	128	53.4	13.4	154765	5	AC114649	AC114649 Mus muscu
C	129	53.4	13.4	154814	5	AC093533	AC093533 Homo sapi
	130	53.4	13.4	163604	14	AC092727	AC092727 Bos tauru
	131	53.4	13.4	170361	6	AC110554	AC110554 Mus muscu
	132	53.4	13.4	177570	11	CR354422	CR354422 Zebrafish
	133	53.4	13.4	187423	12	AC034175	AC034175 Homo sapi
	134	53.4	13.4	187487	6	AC159969	AC159969 Mus muscu
	135	53.4	13.4	195433	6	AC112925	AC112925 Mus muscu
	136	53.4	13.4	213461	12	AC182460	AC182460 Gallus ga
	137	53.4	13.4	217606	6	AC113176	AC113176 Mus muscu
	138	53.4	13.4	237754	12	AC117125	AC117125 Rattus no
C	139	53.4	13.4	239559	12	AC095531	AC095531 Rattus no
	140	53.4	13.4	249062	12	AC160319	AC160319 Bos tauru
	141	53.4	13.4	250870	12	AC177806	AC177806 Gallus ga
	142	53.4	13.4	283738	12	AC178989	AC178989 Strongylo
	143	53.2	13.3	388	7	H8168XC3	Z67781 H. sapiens D
	144	53.2	13.3	113252	12	CT027769	Danio rer
	145	53.2	13.3	116250	6	AC153619	AC153619 Mus muscu
	146	53.2	13.3	140087	12	CR931815	Danio rer
	147	53.2	13.3	149596	12	AC009718	Mus muscu
C	148	53.2	13.3	154123	5	AC138058	AC138058 Homo sapi
C	149	53.2	13.3	155023	11	CR354440	CR354440 Zebrafish
	150	53.2	13.3	158636	12	CR450843	CR450843 Danio rer

## ALIGNMENTS

RESULT 1  
 AYE75320/c 32745 bp DNA linear PRI 17-JUL-2004  
 LOCUS Homo sapiens RAD21 homolog (S. pombe) (RAD21) gene, complete cds.  
 DEFINITION AYE75320  
 ACCESSION AYE75320  
 VERSION AYE75320.1 GI:50234988  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE	1 (bases 1 to 32745)	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	Livingston,R.J., Rieder,M.J., Chung,M.-W., Ritchie,T.K., Olson,A.N., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Chambers,S.W., Schackwitz,W.S., Sherwood,J.K., Sherwood,A.M., Leitnauser,B.J. and Nickerson,D.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-JUL-2004) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA	
COMMENT	To cite this work please use: NIHES-SNPs, Environmental Genome Project, NIHES ES15478, Department of Genome Sciences, Seattle, WA (URL: <a href="http://sepp.gs.washington.edu">http://sepp.gs.washington.edu</a> ).	
FEATURES	location/Qualifiers	
source	1..32745	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
repeat_region	1..1278	
		/rpt_family="T1"
variation	99	/rpt_type="dispersed"
		/frequency="0.02"
variation	120	/replace="g"
		/frequency="0.18"
variation	688	/replace="c"
		/frequency="0.03"
variation	846	/replace="c"
		/frequency="0.24"
variation	1016	/replace="a"
		/frequency="0.01"
variation	1062	/replace="c"
		/frequency="0.06"
variation	1179	/replace="a"
		/frequency="0.01"
variation	1248	/replace="g"
		/frequency="0.03"
variation	1338	/replace="c"
		/frequency="0.24"
variation	1350	/replace="a"
		/frequency="0.01"
variation	1422	/replace="g"
		/frequency="0.06"
variation	1464	/replace="c"
		/frequency="0.02"
variation	1815	/replace="c"
		/frequency="0.01"
variation	1817	/replace="c"
		/frequency="0.15"
variation	1946	/replace="g"
		/frequency="0.01"
gene	join(1946..2097,9949..10124,13451..13580,14770..14869,18253..18359,19238..19444,19940..20065,20423..20545,22243..22466,24003..24162,24615..24763,25944..26093,27682..27765,29020..30782)	
		/gene="RAD21"
		/product="RAD21 homolog (S. pombe)"
variation	2079	







Db	4027	ATGACGTTTCACTTTCCCTACTCTATGAAAACTGGAAAGTTGAAACTGTGCTTTACTCTAA	3968
07	361	CCTTGAAAAAGTCAATACCTTACTCAGATCAGAACGAG	400
Db	3967	CCTTGAAAAAGTCAATACCTTACTCAGATCAGAACGAG	3928
RESULT 2			
AC104391/c			
LOCUS	AC104391	148386 bp	DNA linear PRI 01-MAY-2002
DEFINITION	Homo sapiens chromosome 8, clone CTD-3071K10, complete sequence.		
ACCESSION	AC104391		
VERSION	AC104391.5	GI:20377041	
KEYWORDS	HTG.		
ORGANISM	Homo sapiens (human)		
LOCUS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
DEFINITION	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
ACCESSION	Homnidae; Homo.		
VERSION	1 (bases 1 to 148386)		
REFERENCE	1	(bases 1 to 148386)	
AUTHORS	Bitren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 8, clone CTD-3071K10		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 148386)		
AUTHORS	Bitren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Baetien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,S., Choepel,Y., Colangelo,M., Collins,S., Collingmore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McNeeters,R., Meldrim,J., Menue,L., Mhova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Noddu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roesti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stenge-Thomann,N., Stojanovic,N., Strauss,N., Subramaniam,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Submitted (08-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
JOURNAL	Research, 1 to 148386		
REFERENCE	3 (bases 1 to 148386)		
AUTHORS	Bitren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collingmore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menue,L., Mhova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Noddu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roesti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stenge-Thomann,N., Stojanovic,N., Strauss,N., Subramaniam,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		

[illegible]



OY 1 AAGATGCTTTCTCTATATTTACGAGTCTTGATAAGCTTAAATACCTGTGGC 60  
 Db 129948 AAGATGCTTTCTCTATATTTACGAGTCTTGATAAGCTTAAATACCTGTGGC 129895

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H_NH03167C15
----- Summary Statistics -----
Sequencing vector: M13; 76%
Sequencing vector: plasmid; 24%
Chemistry: Dye-primer ET; 76% of reads
Chemistry: Dye-terminator Big Dye; 24% of reads
Assembly program: Phrap; version 0.990019
Consensus quality: 186568 bases at least Q40
Consensus quality: 189080 bases at least Q30
Consensus quality: 190451 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 192300; sum-of-consigs
Quality coverage: 4.33 in Q20 bases; agarose-fp
Quality coverage: 4.86 in Q20 bases; sum-of-consigs
-----

```



```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1597: contig of 1597 bp in length
*      .1598      1697: gap of unknown length
*      1698      8306: contig of 6609 bp in length
*      8307      8406: gap of unknown length
*      8407      15335: contig of 6929 bp in length
*      15336      15435: gap of unknown length
*      15436      22120: contig of 6685 bp in length
*      22121      22220: gap of unknown length
*      22221      30889: contig of 8669 bp in length
*      30890      30989: gap of unknown length
*      30990      39394: contig of 8405 bp in length
*      39395      39494: gap of unknown length
*      39495      56109: contig of 16615 bp in length
*      56110      56209: gap of unknown length
*      56210      99245: contig of 43036 bp in length
*      99246      99245: gap of unknown length
*      99346      193300: contig of 93755 bp in length.
*
* Location/Qualifiers

```

```

source
1. .193100
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-367C15"
1. .1597
/note="assembly_name:Contig3"
1598. .1697
/estimated_length=unknown
1698. .8306
/note="assembly_name:Contig4"
8307. .8406
/estimated_length=unknown
8407. .15335
/note="assembly_name:Contig5"
15336. .15435
/estimated_length=unknown
15436. .22120
/note="assembly_name:Contig6"
clone_end:r7
vector_side:right"
22121. .22220
/estimated_length=unknown
22221. .30889
/note="assembly_name:Contig7"
30890. .30989
/estimated_length=unknown
30990. .39394
/note="assembly_name:Contig8"
39395. .39494
/estimated_length=unknown
39495. .56109
/note="assembly_name:Contig9"
56110. .56209
/estimated_length=unknown
56210. .99245
/note="assembly_name:Contig10"
99246. .99345
/estimated_length=unknown
99346. .193100
/note="assembly_name:Contig11"
clone_end:SP6
vector_side:left"

```

ORIGIN	
Query Match	99.6%; Score 397.6; DB 12; Length 193100;
Best Local Similarity	99.2%; Pred. 2.8e-123;
Matches 397; Conservative	2; Mismatches 0; Indels 0; Gaps 0;

QY	1	AAGATGCTCTTTCCTATATATTTAAACGAGTCTTGAATPAACCTPAAAATACCTGGGCG	60
Db	192520	AAGAAATGCTCTTTCCTATATATTTAAACGAGTCTTGAATPAACCTPAAAATACCTGGGCG	192461
QY	61	TACACATGAGACTGSGTAGTACTGTTACCTCTGGAAGTAGTATGACAGGCAAGTAGG	120
Db	192460	TACACATGAGACTGSGTAGTACTGTTACCTCTGGAAGTAGTATGACAGGCAAGTAGG	192401
QY	121	CCTACCTTTTGTACTGTTGATATATATATATATACATACACACACACACACACA	180
Db	192400	CCTACCTTTTGTACTGTTGATATATATATATATATACATACACACACACACACACA	192341
QY	181	CAVATATTTTAACTGGGACATAAATGTACACACCTATTCGAAAGTAAAACTACTATCT	240
Db	192340	CACATATTTTAACTGGGACATAAATGTACACACCTATTCGAAAGTAAAACTACTATCT	192281
QY	241	CAGAAGTAAAGACTATATCTGCTATATGTCAATTAATAGCAGTTAAAGTGCCTAACCTT	300
Db	192280	CAGAAGTAAAGACTATATCTGCTATATGTCAATTAATAGCAGTTAAAGTGCCTAACCTT	192221
QY	301	ATGCAAGTTCAGTTTCCCTACATATGAAAACTGAAAGGTTGAAATCTGGTGTTTACCTAA	360
Db	192220	ATGCAAGTTCAGTTTCCCTACATATGAAAACTGAAAGGTTGAAATCTGGTGTTTACCTAA	192161
QY	361	CCTTTGAAAAAGCTCAATACCTAATCTCAGATCGAAGCAGGAG	400
Db	192160	CCTTTGAAAAAGCTCAATACCTAATCTCAGATCGAAGCAGGAG	192121

```

ORIGIN
CS200151      193363 bp      DNA      linear      PAT 08-DEC-20
LOCUS          CS200151
DEFINITION     Sequence 32 from Patent WO2005106044.
ACCESSION      CS200151
VERSION        CS200151.1  GI:83408556
KEYWORDS
SOURCE
ORGANISM       Homo sapiens (human)
                Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homidae; Homo.
REFERENCE
1
  AUTHORS       Harris, C.
  JOURNAL       Patent: WO 2005106044-A 32 10-NOV-2005;
                Exagen Diagnostics, Inc. (US)
FEATURES
  source
    1..193363
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"

```

Query Match	Similarity	99.6%	Score 397.6	DB 2	Length 193363
Best Local	Similarity	99.2%	Pred. No. 2.8e-123		
Matches 397	Conservative	2	Mismatches 1	Indels 0	Gaps 0
Qy	1	AAGATGCTCTTCTCTATATATTTACGAGTTCCTGAATACCTTAAATTAACCTGTGGCG	60		
Db	100161	AAGAAAGCTCTTCTCTATATATTTACGAGTTCCTGAATACCTTAAATTAACCTGTGGCG	100220		
Qy	61	TACACATGACGTGSGATGACTGTTACCTCTCGAAAGTAGGATGCAGGGCAAGTAGG	120		
Db	100221	TACACATGAACTGGTGGATCTGGTTACCTCTCGAAAGTAGGATGCAGGGCAAGTAGG	100280		
Qy	121	CCTACCTTTTGTACTGTTGATATATATATATATACACACACACACACACACAC	180		
Db	100281	CCTACCTTTTGTACTGTTGATATATATATATATACACACACACACACACACAC	100340		
Qy	181	CAVATATTTTAACGCGGACATTAATAGTACACACCTATTCAAAGTAAATCACTATCT	240		
Db	100341	CACATATTTTAACGCGGACATTAATAGTACACACCTATTCAAAGTAAATCACTATCT	100400		
Qy	241	CAGAACTAAAGCATATATCTCGCTATGTGCTATTAACCTAGCACTTAAGCTGCTT	300		



	DB	100401	CAGAACTAAGACATCATTCCTGCTATTGTGCATTACTAGCAGTTAACGCTCCATACTTT	100466
	OY	301	ATGCAGTTCAGTTTCCCTACATATGAAAACCTGAAGGTTGAAATCTGGTGTTTACCCTAA	360
	DB	100461	ATGCAGTTCAGTTTCCCTACTACTGAAAACTGTAAAGSTTGAAATCTGGTGTTTACCCTAA	100520
	OY	361	CCTTTGAAAAACGCTCAATACTACTACTGAGTTCAGAACGAGG	400
	DB	100521	CCTTTGAAAAACGCTCAATACTACTACTGAGTTCAGAACGAGG	100560
	RESULT 5			
	LOCUS	AC087350/c		
	DEFINITION	Homo sapiens chromosome 8, clone RP11-367C15, complete sequence.		
	ACCESSION	AC087350		
	VERSION	AC087350.3	GI:22267854	
	KEYWORDS	HTG.		
	SOURCE	Homo sapiens (human)		
	ORGANISM	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
		Homnidae; Homo.		
	REFERENCE	1 (bases 1 to 193364)		
	AUTHORS	Birren,B., Nusbaum,C. and Lande,E.		
	TITLE	Homo sapiens chromosome 8, clone RP11-367C15		
	JOURNAL	Unpublished		
	REFERENCE	2 (bases 1 to 193364)		
	AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lande,E., Allen,N., Anderson,S.,		
		Barna,N., Baetien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,		
		Camrata,U., Campiano,A., Choepel,Y., Colangelo,M., Collins,S.,		
		Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,		
		Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,		
		Gardyn,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,		
		Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,		
		Jones,C., Karacas,A., Labocque,K., Lamazares,R., Landers,T.,		
		Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P.,		
		Marquis,N., Matthews,C., McCarthy,M., Mcwan,P., McKernan,K.,		
		McPheeters,R., Meldrum,J., Meneis,L., Mihova,T., Mlenga,V.,		
		Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,		
		O'Connor,T., O'Donnell,E., O'Neill,D., Oliver,J., Peterson,K.,		
		Pinhuangang,P., Pletre,N., Pollara,V., Raymond,C., Retta,R.,		
		Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselti,M.,		
		Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,		
		Souner,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,		
		Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,		
		Trevers,N., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,		
		Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,		
		Zembek,L., Zimmer,A. and Zody,M.		
	TITLE	Direct Submission		
	JOURNAL	Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome		
	REFERENCE	Research, 320 Charles Street, Cambridge, MA 02141, USA		
	AUTHORS	3 (baes 1 to 193364)		
		Birren,B., Nusbaum,C., Lande,E., Ali,A., Allen,N., Anderson,S.,		
		Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,		
		Camaratat,J., Chang,J., Chazarro,B., Choepel,Y., Collymore,A.,		
		Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,		
		Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,		
		Gardyna,S., Gord,S., Graham,S., Grand-Pierre,N., Hagos,B.,		
		Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,		
		Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,		
		Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,		
		McCarthy,M., Meldrum,J., Meneis,L., Mihova,T., Mlenga,V.,		
		Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,		
		O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,		
		Pinhuangang,P., Pletre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,		
		Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,		
		Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,		
		Tesfaye,S., Theodore,J., Topham,K., Travas,M., Vassiliev,H.,		
		Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,		
		Zembek,L., Zimmer,A. and Zody,M.		
		Direct Submission		

JOURNAL	Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	4 (bases 1 to 193364)
AUTHORS	Bairn, N., Nusbach, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barrin, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gadyra, S., Gord, S., Graham, L., Grand-Pierre, N., Hagob, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menais, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Punthang, P., Pierre, N., Raymond, C., Retta, R., Rhee, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Rhoman, N., Stojanovic, N., Talama, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Aug 16, 2002 this sequence version replaced gi:16905274. All repeats were identified using RepeatMasker: Smith, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/BM/RepeatMasker.html">http://ftp.genome.washington.edu/BM/RepeatMasker.html</a> ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence.submissions@genome.wi.mit.edu">sequence.submissions@genome.wi.mit.edu</a> ----- Project Information Center project name: L1698 Center clone name: 367_C15 ----- Location/Qualifiers 1. 193364 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="8" /map="8" /clone="RP11-367C15" /clone_1fb="RP11-11 Human Male BAC" 290..348 /rpt_family="MIR" 349..651 /rpt_family="AluSg" 652..730 /rpt_family="MIR" 1721..2290 /rpt_family="L1MB5" 3577..3711 /rpt_family="MSTD" 3728..3786 /rpt_family="PRAM" 3812..4016 /rpt_family="MSTD" complement(4135..4438) /rpt_family="AluY" 4531..4597 /rpt_family="(TTCC)n" 5406..6437 /rpt_family="L1MBC" 6438..6744 /rpt_family="AluY" 6745..7328 /rpt_family="L1MBC" 8016..8321 /rpt_family="AluX" complement(8336..8431)



repeat_region	/rpt_family="L13"	complement(9625, .9835)
repeat_region	/rpt_family="MULTIC"	complement(9836, .10138)
repeat_region	/rpt_family="AluSx"	complement(10139, .10431)
repeat_region	/rpt_family="MULTIC"	10441, .11704
repeat_region	/rpt_family="L2"	12286, .12710
repeat_region	/rpt_family="L1PA16"	12715, .12735
repeat_region	/rpt_family="("AAA)n"	12805, .12829
repeat_region	/rpt_family="HML1"	12830, .13136
repeat_region	/rpt_family="AluSx"	13137, .13265
repeat_region	/rpt_family="HML1"	13791, .13643
repeat_region	/rpt_family="L2"	complement(14375, .14530)
repeat_region	/rpt_family="MIR3"	15285, .16638
repeat_region	/rpt_family="PTR5"	complement(16805, .17197)
repeat_region	/rpt_family="MULTIL"	complement(17353, .17462)
repeat_region	/rpt_family="MIR"	complement(17890, .18020)
repeat_region	/rpt_family="MIR3"	18233, .18405
repeat_region	/rpt_family="MIR"	complement(18743, .18834)
repeat_region	/rpt_family="MIR"	18857, .18915
repeat_region	/rpt_family="MIR"	complement(21125, .21435)
repeat_region	/rpt_family="AluSx"	21493, .21528
repeat_region	/rpt_family="AT_rich"	complement(22565, .22867)
repeat_region	/rpt_family="AluY"	23040, .23060
repeat_region	/rpt_family="AT_rich"	24278, .24530
repeat_region	/rpt_family="L2"	24532, .24562
repeat_region	/rpt_family="("TTAA)n"	complement(24678, .24746)
repeat_region	/rpt_family="Charlie4"	complement(25284, .25429)
repeat_region	/rpt_family="MIR"	25489, .25526
repeat_region	/rpt_family="5S"	complement(25586, .25737)
repeat_region	/rpt_family="MIR3"	complement(25941, .26235)
repeat_region	/rpt_family="AluSq"	27079, .27100
repeat_region	/rpt_family="("CCG)n"	28056, .28463
repeat_region	/rpt_family="MSRA"	28644, .28671
repeat_region	/rpt_family="("TTCA)n"	complement(28724, .28949)
repeat_region	/rpt_family="MIR"	complement(29453, .29575)
Query Match	99.64;	Score 397.6; DB 5;
Best Local Similarity	99.24;	Pred. No. 2.8e-123;
Matches 397; Conservative	2;	Mismatches 1;
		Indels 0; Gaps 0

OY	1	AAGAAAGCTCTTCTCTATATATTTACGAGTCTTGAATPAAGTTAAATACCTGGAGG	60
Db	93203	AAGAAATCTTCTCTATATATTTACGAGTCTTGAATPAAGTTAAATACCTGGAGG	93144
OY	61	TACACATGAGACGTGATCTGTTACCTCTGGAAGTAGTATCAGGGCAAGTAG	120
Db	93143	TACACATGAAACGTGATCTGTTACCTCTGGAAGTAGTATCAGGGCAAGTAG	93088
OY	121	CTTACCTTTTGTACTGTTGATATATATATATACACACACACACACACACA	180
Db	93083	CTTACCTTTTGTACTGTTGATATATATATATACACACACACACACACACACA	93024
OY	181	CAVATATTTTAACTCGGGACATTAAGTACACACCATTCAAAGTAAATACATCT	240
Db	93023	CACATATTTTAACTCGGGACATTAAGTACACACCATTCAAAGTAAATACATCT	92964
OY	241	CAGAAGTAAAGACATAATCTCGTATATGTCATTAACTAGCAGTTAAGTCCTAAC	300
Db	92963	CAGAAGTAAAGACATAATCTCGTATATGTCATTAACTAGCAGTTAAGTCCTAAC	92904
OY	301	ATGCAGTTCAAGTTCCCTACTATGAAAACTGAAAGGTTGAAATCTGGTGTTTAC	360
Db	92903	ATGCAGTTCAAGTTCCCTACTATGAAAACTGAAAGGTTGAAATCTGGTGTTTAC	92844
OY	361	CCTTGAAGAAAGTCAATACCTAATCTAGATCAAGCAAGGAG	400
Db	92843	CCTTGAAGAAAGTCAATACCTAATCTAGATCAAGCAAGGAG	92804

```

RESULT 6
AP004074      156441 bp      DNA      linear      PRI 21-AUG-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..156441
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q23"
/clone="KB1266F9"
/cell_line="HEB 14-14"
/clone_1lb="Keio BAC library"
560..581
/inference="non-experimental evidence, no additional
details recorded"
723..3033
/rpc_family="AT_rich"
/inference="non-experimental evidence, no additional
details recorded"
/rpc_family="WSTB"
3626..3688
/inference="non-experimental evidence, no additional

```



repeat_region	details recorded" /rpt_family="A)n" 5506..5537 /inference="non-experimental evidence, no additional details recorded" /rpt_family="CT-rich" 7579..7632 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="Tigger1" 14872..15595 /inference="non-experimental evidence, no additional details recorded" /rpt_family="Tigger1" 15637..15922 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" complement(7956..8247) /inference="non-experimental evidence, no additional details recorded" /rpt_family="A1usx" 8697..8731 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MC5" 17431..17469 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="L2" 9452..9560 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="AT-rich" 19504..19526 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" 10358..10385 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MC1" complement(22822..23099) /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" 10421..10441 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="A1uvo" complement(23100..23442) /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" complement(10443..10592) /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MD1" 23442..24177 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="FLAM_C" 11163..11185 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MC/D" 24178..24474 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="A)n" 11849..11876 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="A1uy" 24475..24685 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" 12695..12861 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MC/D" 25120..25255 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="FAM" 13403..13423 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="FLAM_C" complement(26608..26781) /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="AT-rich" 13609..13719 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="MIR" complement(26999..27171) /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="MIR" 13731..13758 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="MER2" 27399..27410 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="L1MC5" 14034..14260 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="AT-rich" 27411..27435 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="Tigger1" complement(14263..14407) /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="CA)n" 29520..29564 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="FLAM_A" 14409..14877 /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="G-rich" 30299..32109 /inference="non-experimental evidence, no additional details recorded"
repeat_region	/rpt_family="L1MA4" complement(14263..14407) /inference="non-experimental evidence, no additional details recorded"	repeat_region	/rpt_family="L1MA4" complement(14263..14407) /inference="non-experimental evidence, no additional details recorded"



[illegible]

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
Unpublished				Penhader, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gablis, A., Gant, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Gladly, M., Guerra, M., Guivar, M., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hernandez, J., Harvey, Y., Havak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, B., Hines, S., Hladun, S., L., Hodgson, A., Hughes, M., Hollins, B., Howell, S., Huliy, K., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Jorkeby, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovis, C., Kraft, C.L., Lebow, A., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzshwa, L., Louised, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahndartine, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapu, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T. Z., Meenan, E., Mitovaljevic, A., Miner, G., Minda, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundassa, M., Murphy, M., Nair, L., Narkavici, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokediemeh, O., Okunou, G., Olarunsegun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poidexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Rul, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorella, R., Sosa, J., Steime, M., Strong, R., Sutton, A., Svatke, A., Taboz, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trebor, Z., Usami, K., Valas, R., Wang, S., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, R., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., Von Niederhausen, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.			
Direct Submission							
2 (bases 1 to 211875)							
Worley, K.C.							
Direct Submission							
Submitted (01-NOV-2004)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
3 (bases 1 to 211875)							
Cow Genome Sequencing Consortium.							
Direct Submission							
Submitted (01-JUL-2005)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
On Jun 26, 2005 this sequence version replaced gi:58038130.							
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.							
----- Genome Center							
Center: Baylor College of Medicine							
Center code: BCM							
Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>							
Contact: hgsc-help@bcm.tmc.edu							
----- Project Information							
Center project name: FABZ							
Center clone name: CH240-1K23							
----- Summary Statistics							
Assembly program: Atlas 3.0;							



Consensus quality: 198612 bases at least Q40  
 Consensus quality: 201789 bases at least Q30  
 Consensus quality: 204577 bases at least Q20  
 Estimated insert size: 203572; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 32 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 30291: contig of 30291 bp in length  
 30292 30341: gap of 50 bp  
 30342 37665: contig of 7324 bp in length  
 37666 37715: gap of 50 bp  
 37716 53826: contig of 16111 bp in length  
 53827 53876: gap of 50 bp  
 53877 56440: contig of 2564 bp in length  
 56441 56490: gap of 50 bp  
 56491 67817: contig of 11327 bp in length  
 67818 68524: gap of 707 bp  
 68525 70609: contig of 2085 bp in length  
 70610 71044: gap of 435 bp  
 71045 73476: contig of 2432 bp in length  
 73477 73526: gap of 50 bp  
 73527 92741: contig of 19215 bp in length  
 92742 92791: gap of 50 bp  
 92792 96778: contig of 3987 bp in length  
 96779 96828: gap of 50 bp  
 96829 105197: contig of 8369 bp in length  
 105198 105297: gap of unknown length  
 105298 106706: contig of 1409 bp in length  
 106707 106756: gap of 50 bp  
 106757 115759: contig of 9003 bp in length  
 115760 115859: gap of unknown length  
 115860 122530: contig of 6671 bp in length  
 122531 122662: gap of 132 bp  
 12663 126722: contig of 4060 bp in length  
 126723 126772: gap of 50 bp  
 126773 131853: contig of 5081 bp in length  
 131854 132386: gap of 533 bp  
 132387 140048: contig of 7662 bp in length  
 140049 140098: gap of 50 bp  
 140099 148892: contig of 8794 bp in length  
 148893 148950: gap of 58 bp  
 148951 159941: contig of 10991 bp in length  
 159942 159991: gap of 50 bp  
 159992 167844: contig of 7853 bp in length  
 167845 167944: gap of unknown length  
 167945 177933: contig of 9989 bp in length  
 177934 178226: gap of 293 bp  
 178227 182916: contig of 4690 bp in length  
 182917 182966: gap of 50 bp  
 182967 185344: contig of 2378 bp in length  
 185345 185394: gap of 50 bp  
 185395 189477: contig of 4083 bp in length  
 189478 189527: gap of 50 bp  
 189528 192243: contig of 2716 bp in length  
 192244 192293: gap of 50 bp  
 192294 195792: contig of 3499 bp in length  
 195793 195842: gap of 50 bp  
 195843 204127: contig of 8285 bp in length  
 204128 204227: gap of unknown length  
 204228 205866: contig of 1639 bp in length  
 205867 205966: gap of unknown length  
 205967 207177: contig of 1211 bp in length  
 207178 207277: gap of unknown length  
 207278 208398: contig of 1121 bp in length

FEATURES  
 SOURCE  
 1..211875  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9913"  
 /clone="CH240-1K23"  
 30292..30341  
 /estimated\_length=50  
 37666..37715  
 /estimated\_length=50  
 53827..53876  
 /estimated\_length=50  
 56441..56490  
 /estimated\_length=50  
 67818..68524  
 /estimated\_length=707  
 70610..71044  
 /estimated\_length=435  
 73477..73526  
 /estimated\_length=50  
 92742..92791  
 /estimated\_length=50  
 96779..96828

Query Match 15.6%; Score 62.2; DB 12; Length 211875;  
 Best Local Similarity 61.3%; Pred. No. 1.9e-09;  
 Matches 100; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 8 CTCCTTCTCTATATTATTTACAGTTCCTGAATAGCTTAATATCTGTGCGTACACAT 67  
 Db 71851 CTCCTATAGTATGTTATTAACGACCTTGCGAGCATTAATATCTGCAATGAATAC 71910  
 Qy 68 GAGACTGTGATGTTACTGTTACTCTGGAAGTATGACGGCAAGTGGAGCTTACTT 127  
 Db 71911 AGAAGTATGATGATGCTGCTTGAAGTATGAGTGGTGAAGGAGCCAGCTTACC 71970  
 Qy 128 TTTTGTACTGTTTAT 170  
 Db 71971 TTTGCACTGACGGCTTTTATTAACATATAGGATATACAA 72013

RESULT 8  
 LOCUS AC122925 256751 bp DNA linear ROD 08-NOV-2003  
 DEFINITION Mus musculus BAC clone RP23-9019 from 1, complete sequence.  
 ACCESSION AC122925  
 VERSION AC122925.3 GI:24476166  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 1 (bases 1 to 256751)  
 AUTHORS Harkins, R., Doeber, A. and Biellack, L.  
 TITLE The sequence of Mus musculus BAC clone RP23-9019  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 256751)  
 AUTHORS Wilson, R.  
 TITLE Sequencing of Mus musculus  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 256751)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 256751)



AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 256751)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-NOV-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 6 (bases 1 to 256751)  
 AUTHORS Wilson, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Nov 2, 2002 this sequence version replaced gi:23334975.  
 COMMENT ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: M\_BA009019  
 -----  
 NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
 MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>  
 SOURCE INFORMATION:  
 The RPCI-23 BAC Library has been constructed by Kazuroyo Oesegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 NEIGHBORING SEQUENCE INFORMATION:  
 This sequence is the entire insert of the clone.  
 Location/Qualifiers  
 1..256751  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="1"  
 /clone="RP23-9019"  
 /clone\_1lb="RPCI-23"  
 repeat\_region 3239..3383  
 /rpt\_family="A1u"  
 repeat\_region 3957..4273  
 /rpt\_family="MALR"  
 repeat\_region 5279..5419  
 /rpt\_family="ERVK"  
 repeat\_region 5429..6070  
 /rpt\_family="L1"  
 repeat\_region 6071..6562  
 /rpt\_family="ERVK"  
 repeat\_region 8382..8633

repeat\_region 8629..8747  
 /rpt\_family="ERVK"  
 repeat\_region 8748..9301  
 /rpt\_family="ERVK"  
 repeat\_region 9302..9467  
 /rpt\_family="ERVK"  
 repeat\_region 9523..10022  
 /rpt\_family="ERVK"  
 repeat\_region 10023..10101  
 /rpt\_family="L1"  
 repeat\_region 13484..13691  
 /rpt\_family="B4"  
 repeat\_region 15335..15475  
 /rpt\_family="B4"  
 repeat\_region 15614..16232  
 /rpt\_family="L1"  
 repeat\_region 17010..17083  
 /rpt\_family="A1u"  
 repeat\_region 18336..18496  
 /rpt\_family="L1"  
 repeat\_region 19428..19496  
 /rpt\_family="ID"  
 repeat\_region 19515..19733  
 /rpt\_family="B4"  
 repeat\_region 20439..20626  
 /rpt\_family="L1"  
 21165..21733  
 /note="Sequence derived from PCR product of genomic DNA."  
 21885..22150  
 /note="Sequence derived from one plasmid subclone."  
 26874..27413  
 /rpt\_family="ERVK"  
 27428..27573  
 /rpt\_family="B4"  
 27933..28326  
 /rpt\_family="MALR"  
 28769..29572  
 /rpt\_family="ERVK"  
 29712..29928  
 /rpt\_family="L1"  
 29928..30367  
 /rpt\_family="L1"  
 30828..30886  
 /rpt\_family="B4"  
 33156..33251  
 /rpt\_family="ERVJ"  
 33401..33512  
 /rpt\_family="L1"  
 34239..34610  
 /rpt\_family="MALR"  
 37270..37700  
 /rpt\_family="L1"  
 38033..38160  
 /rpt\_family="B4"  
 38129..38179  
 /rpt\_family="B4"  
 38681..38886  
 /rpt\_family="L1"  
 39047..39479  
 /rpt\_family="EMER6A"  
 39718..40075  
 /rpt\_family="L1"  
 40191..40305  
 /rpt\_family="L2"  
 42041..42191  
 /rpt\_family="A1u"  
 42054..42228  
 /rpt\_family="B4"  
 42960..43145  
 /rpt\_family="L1"  
 47148..47311  
 /rpt\_family="MIR"



Query Match	15.2%	Score 60.8;	DB 6;	Length 256751;
Best Local Similarity	63.0%;	Pred. No. 5.8e-09;		
Matches 92; Conservative	1;	Mismatches 53;	Indels 0;	Gaps 0;

## RESULT 9

## ORGANISM

**COMMENT**

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

1

Matches 108

D**b**[illegible]











```

repeat_region      /rpt_family="B2_Mm2"
                    3830..3893
repeat_region      /rpt_family="Lx7"
                    complement(3894..4094)
repeat_region      /rpt_family="ORR1B2"
                    4095..4286
repeat_region      /rpt_family="B2_Mm1"
                    complement(4287..4453)
repeat_region      /rpt_family="ORR1B2"
                    4454..4536
repeat_region      /rpt_family="Lx7"
                    complement(4537..4786)
repeat_region      /rpt_family="ORR1B1"
                    4828..5067
repeat_region      /rpt_family="Lx7"
                    5068..5100
repeat_region      /rpt_family="TTTG)n"
                    5242..5438
repeat_region      /rpt_family="B2_Mm1"
                    5506..5643
repeat_region      /rpt_family="MTD"
                    5773..5862
repeat_region      /rpt_family="MLT2B3"
                    5887..5950
repeat_region      /rpt_family="(TG)n"
                    complement(5979..6235)
repeat_region      /rpt_family="B4A"
                    complement(6276..6412)
repeat_region      /rpt_family="B1_MM"
                    6821..6968
repeat_region      /rpt_family="B1_MM"
                    6969..6991
repeat_region      /rpt_family="(GAAA)n"
                    7195..7247
repeat_region      /rpt_family="(TACA)n"
                    complement(7319..7589)
repeat_region      /rpt_family="MLT2B1"
                    complement(7672..8158)
repeat_region      /rpt_family="RLTR11B"
                    8429..9003
repeat_region      /rpt_family="L1MB8"
                    complement(9067..9412)
repeat_region      /rpt_family="RLTR11B"
                    complement(9630..9829)
repeat_region      /rpt_family="B3A"
                    10359..10587
repeat_region      /rpt_family="(TTTTTG)n"
                    complement(10588..10738)
repeat_region      /rpt_family="B1_MM"
                    complement(11080..11165)
repeat_region

Query Match      14.94; Score 59.6; DB 6; Length 181457;
Best Local Similarity 58.04; Pred. No. 1.4e-08;
Matches 101; Conservative 2; Mismatches 71; Indels 0; Gaps 0;

QY 35 TGAATAGCTTAATAATCGTGGCGTACATAGAGCTGTGATCTGTTGATATATATA 154
DB 79483 TCAATAGTGCACAGTGTGGGCTCTTGGATGGGATTAATTTTTCCTTTTA 79542
QY 95 AAATAGGTATGAGGGCGAGTGAAGGCTTACCTTTTGTCTGTTGATATATATA 154
DB 79543 TGGTTTGATTAATAAAGATGATGTTGTTGTTAAATCTTTTGGATACATACACA 79602
QY 155 CATACACACACACACACACACACACACATATTTTAACTGGGACATAATGT 208
DB 79603 CACACACACACACACACACACACACACACACACACAGCCAGATGT 79656

```

RESULT 12  
AC130372 AC130372 64239 bp DNA linear HTG 30-AUG-2002  
LOCUS Homo sapiens chromosome 11 clone RP11-1360C17 map 11, LOW-PASS  
DEFINITION SEQUENCE SAMPLING.

```

ACCESSION      AC130372
VERSION        AC130372.2
KEYWORDS      HTGS_PHASE50
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 64239)
AUTHORS        Birren,B., Nusbaum,C. and Lander,E.
TITLE          Homo sapiens chromosome 11, clone RP11-1360C17
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 64239)
AUTHORS        Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
                Barre,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
                Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
                Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
                Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
                Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
                Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
                Karakas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
                Liu,G., Maclean,C., Macdonald,P., Major,U., Matthews,C.,
                McCarthy,M., Meldrum,J., Meneus,L., Minova,T., Mienga,V.,
                Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
                O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
                Plunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
                Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
                Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
                Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
                Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
                Zembek,L., Zimerer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE      3 (bases 1 to 64239)
AUTHORS        Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
                Barre,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
                Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
                Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
                Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
                Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
                Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
                Karakas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
                Liu,G., Maclean,C., Macdonald,P., Major,U., Matthews,C.,
                McCarthy,M., Meldrum,J., Meneus,L., Minova,T., Mienga,V.,
                Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
                O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
                Plunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
                Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
                Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
                Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
                Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
                Zembek,L., Zimerer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On Aug 30, 2002 this sequence version replaced gi:22165307.
                All repeats were identified using RepeatMasker:
                http://ftp.genome.washington.edu/RM/RepeatMasker.html
                ----- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WtBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                ----- Project Information
                Center project name: L27982
                Center clone name: 1360_C_17

```

\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely







[illegible]

```

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 104485)
Worley,K.C.
Direct Submission
Submitted (01-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2002 this sequence version replaced gi:21327308.
COMMENT
gc-help@bcm.tmc.edu
http://www.hgsc.bcm.tmc.edu/ or email

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8086/quality.info/genbank.annotation.ht
ml.

FEATURES
source
Location/Qualifiers
1..104485
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-12K11"
complement(5..2005)
/note="overlaps bases 1..2001 of clone AC128686"
/misc_feature
repeat_region
complement(78..175)
/rpt_family="MER41B"
complement(173..574)
/rpt_family="MER41A"
593..635
/rpt_family="AT_rich"
698..838
/rpt_family="TA1"
839..1149
/rpt_family="ATAT"
1150..1271
/rpt_family="TA1"
1277..1490
/rpt_family="TA1"
complement(1750..1792)
/rpt_family="TA1"
1793..2096

```



```
repeat_region /rpt family="AluSx"
complement(2097..3059)
/rpt family="L1MA10"
3063..3106
/rpt family="L1TTC)n"
complement(3107..3369)
/rpt family="L1MA10"
repeat_region 3963..4232
/rpt family="L1ME3"
repeat_region 4244..4648
/rpt family="L1ME3"
4660..4738
/rpt family="MER21B"
4762..5510
/rpt family="MER21B"
5549..5749
/rpt family="L1ME3"
complement(5904..6104)
/rpt family="MER30"
6723..7134
/rpt family="L1M4"
complement(7682..7854)
/rpt family="L2"
8018..8336
/rpt family="L2"
8356..8549
/rpt family="L2"
complement(8566..8853)
/rpt family="AluSg"
9161..9183
/rpt family="AT_rich"
10173..10233
/rpt family="AluDo"
10234..10939
/rpt family="CTR5"
10941..11678
/rpt family="LTR12"
11679..11769
/rpt family="AluDo"
11770..11927
/rpt family="AluDo"
complement(13036..13368)
/rpt family="L1PA10"
complement(13369..13666)
/rpt family="AluSg"
complement(13667..14581)
/rpt family="L1PA10"
complement(147482..17866)
/rpt family="L1PA12"
18512..18540
/rpt family="AT_rich"
complement(18956..19228)
/rpt family="L1MC/D"
complement(19422..19479)
/rpt family="L1MC/D"
complement(20421..20657)
```

Query Match	14.7%	Score 58.8;	DB 5;	Length 104485;
Best Local Similarity	67.5%	Pred. No. 2.4e-08;		
Matches 81; Conservative	1;	Mismatches 38;	Indels 0;	Gaps 0;

[illegible]

RESULT 14	
AC106720	
LOCUS	
AC106720	92354 bp
	DNA
	linear
	PRI 01-FEB-2003

DEFINITION	Homo sapiens 3 BAC RP11-26k20 (Roswell Park Cancer Institute Human BAC library) complete sequence.
ACCESSION	AC106720
VERSION	AC106720.8
KEYWORDS	GI:22003976
HTG.	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE AUTHORS	1 (bases 1 to 92354)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,	

Aladrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbata, J., Benton, J., Blange, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Boyle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrill, K.L., Byrd, N.C.,  
Carron, T.P., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Dayla, M.L., Davis, C., Davy-Carroll, L., Dedech, D.A.,  
Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dungan-Rocha, S., Dublin, K.U.,  
Ezrahnat, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabalis, A., Gao, Y., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
Hollins, B., Homsl, F., Howard, S., Huber, U., Huiyk, S., Hume, J.,

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 92354)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 92354)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 92354)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	5 (bases 1 to 92354)







LOCUS AC027755 147328 bp DNA linear HTG 02-MAY-2000  
 DEFINITION Homo sapiens chromosome 3 clone RP11-594C8 map 3, WORKING DRAFT  
 ACCESSION AC027755  
 VERSION AC027755.2 GI:7677887  
 KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 147328)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Homo sapiens chromosome 3, clone RP11-594C8  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 147328)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,  
 Campiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karacas, A.,  
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lenoczky, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
 McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R.,  
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenaga, V., Morrow, J.,  
 Murphy, T., Naylor, T. M., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Olivar, T. M., Oliver, J., Peterson, K., Pierre, N., Piant, C.,  
 Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Teefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 JOURNAL Direct Submission  
 COMMENT Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 2, 2000 this sequence version replaced gi:7382581.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

FEATURES  
 source 1..147328  
 location/Qualifiers  
 1..147328  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3"  
 /clone="RP11-594C8"  
 /clone\_lib="RPCT-11 Human Male BAC"  
 1..674  
 /note="assembly\_fragment  
 clone end:77  
 vector side:right"  
 675..774  
 /estimated\_length=100  
 775..3066  
 /note="assembly\_fragment"  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /note="assembly\_fragment"  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /note="assembly\_fragment"  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /note="assembly\_fragment"  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /note="assembly\_fragment"  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /note="assembly\_fragment"  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /note="assembly\_fragment"  
 39993..48004  
 /note="assembly\_fragment"  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=100  
 11453..17821  
 /estimated\_length=100  
 17822..17921  
 /estimated\_length=100  
 17922..25702  
 /estimated\_length=100  
 25703..25802  
 /estimated\_length=100  
 25803..32656  
 /estimated\_length=100  
 32657..32757  
 /estimated\_length=100  
 32758..39992  
 /estimated\_length=100  
 39993..48004  
 /estimated\_length=100  
 48005..48104  
 /estimated\_length=100  
 48105..56776  
 /estimated\_length=100  
 56777..68565  
 /estimated\_length=100  
 68566..68665  
 /estimated\_length=100  
 68666..81186  
 /estimated\_length=100  
 81187..81287  
 /estimated\_length=100  
 81288..97090  
 /estimated\_length=100  
 97091..97191  
 /estimated\_length=100  
 97192..119119  
 /estimated\_length=100  
 119120..147328  
 /estimated\_length=100  
 147329..28109  
 /estimated\_length=100  
 28110..675  
 /estimated\_length=100  
 676..774  
 /estimated\_length=100  
 775..3066  
 /estimated\_length=100  
 3067..3166  
 /estimated\_length=100  
 3167..6613  
 /estimated\_length=100  
 6614..6713  
 /estimated\_length=100  
 6714..11352  
 /estimated\_length=100  
 11353..11452  
 /estimated\_length=10







```

misc_feature      1..1056
                  /note="assembly_fragment"
gap               1057..1156
                  /estimated_length=100
misc_feature      1157..14629
                  /note="assembly_fragment"
gap               4630..4729
                  /estimated_length=100
misc_feature      4730..5865
                  /note="assembly_fragment"
gap               5866..5965
                  /estimated_length=100
misc_feature      5966..9176
                  /note="assembly_fragment"
gap               9177..9276
                  /estimated_length=100
misc_feature      9277..12888
                  /note="assembly_fragment"
gap               12889..12988
                  /estimated_length=100
misc_feature      12989..15553
                  /note="assembly_fragment"
gap               15954..16053
                  /estimated_length=100
misc_feature      16054..21676
                  /note="assembly_fragment"
gap               21677..21776
                  /estimated_length=100
misc_feature      21777..27872
                  /note="assembly_fragment"
gap               27873..27972
                  /estimated_length=100
misc_feature      27973..37824
                  /note="assembly_fragment"
gap               37825..37924
                  /estimated_length=100
misc_feature      37925..47260
                  /note="assembly_fragment"
gap               47261..47360
                  /estimated_length=100
misc_feature      47361..60391
                  /note="assembly_fragment"
gap               60392..60491
                  /estimated_length=100
misc_feature      60492..84436
                  /note="assembly_fragment"
gap               84437..84536
                  /estimated_length=100
misc_feature      84537..114093
                  /note="assembly_fragment"
gap               114094..114193
                  /estimated_length=100
misc_feature      114194..148851
                  /note="assembly_fragment"
ORIGIN
Query Match      14.7%; Score 58.6; DB 12; Length 148851;
Best Local Similarity 57.5%; Pred. No. 3e-08;
Matches 103; Conservative 1; Mismatches 75; Indels 0; Gaps 0;

```

```

Db               27024 TATACATATATATGATATATATACATTTAGCGCTGAAGTAAAGAACTGTATAGACATCAGC 26966
RESULT 17
AC108728/c
LOCUS
DEFINITION
AC108728      303189 bp      DNA      linear      HTG 09-MAY-2002
Homo sapiens chromosome 3 clone RP11-594C8, WORKING DRAFT SEQUENCE,
18 unordered pieces.
ACCESSION
AC108728
VERSION
AC108728.2 GI:20335976
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 303189)
Muzny D.M., Adams C., Adio-Oduola B., Alf-Osman F.R., Allen C.,
Albrooks S.L., Amaralunge H.C., Are J.R., Ayele M., Banks T.,
Barbarta J., Benton J., Binahe K., Blankenburg K., Bonnin D.,
Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,
Bunay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhury I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Datborne S.R., David R.,
Dayila M.L., Davis C., Davy-Carroll J., Dedrich D.A.,
DeJanelle K.R., Delgado O., Denn A.L., Ding Y., Din H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Dumbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Haylak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Homai F., Howard S., Huber J., Hulys K., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsson E., Kelly S., Khan U., Kling L., Korval J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtenage O., Lieu C., Liu J., Liu W., Louised H.,
Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapa P., Martin R., Martindale A., Martinez E.,
Massey G., Mashney E., McLeod M.P., Meador M., Mel G., Metzger M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser M., Neal D., Newton S.J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokkenko S., Ogih M., Okunomu G.,
Oregon N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojokan I., Rolfe M., Ruiz S., Savery G.,
Scheier S., Scott G., Shen H., Shoostari N., Sisson I.,
Sodergren E., Sonaite T., Sparks H., Stanley H., Stone H.,
Sutton A., Svatek A., Taber P., Tameisla A., Tameisla K., Tang H.,
Tansley J., Taylor C., Taylor T., Telford D., Thomas N., Thomas S.,
Umani K., Vasquez L., Vera V., Villalon B., Vinson R., Wang O.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wleczek R., Wooden S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zortilla S., Nelson D.,
Weinstock G. and Gibbs R.
Direct Submission
Unpublished
2 (bases 1 to 303189)
Worley K.C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 303189)
Worley K.C.
Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Apr 28, 2002 this sequence version replaced gi:18449858.

```



```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDVG
Center clone name: RP11-594C8
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 295424 bases at least Q40
Consensus quality: 300495 bases at least Q30
Consensus quality: 303367 bases at least Q20
Estimated insert size: 152199; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3465: contig of 3465 bp in length
* 3466 3565: gap of unknown length
* 3566 7534: contig of 3969 bp in length
* 7535 7634: gap of unknown length
* 7635 13994: contig of 6360 bp in length
* 13995 14094: gap of unknown length
* 14095 22000: contig of 7906 bp in length
* 22001 22101: gap of unknown length
* 22101 30722: contig of 8622 bp in length
* 30723 30823: gap of unknown length
* 30823 42634: contig of 11812 bp in length
* 42635 42734: gap of unknown length
* 42735 54975: contig of 12241 bp in length
* 54976 55075: gap of unknown length
* 55076 67039: contig of 11964 bp in length
* 67040 67140: gap of unknown length
* 67140 79747: contig of 12608 bp in length
* 79748 79847: gap of unknown length
* 79848 91971: contig of 12124 bp in length
* 91972 92071: gap of unknown length
* 92071 108785: contig of 16714 bp in length
* 108786 108885: gap of unknown length
* 108886 130010: contig of 21125 bp in length
* 130011 130110: gap of unknown length
* 130111 146359: contig of 16249 bp in length
* 146360 146459: gap of unknown length
* 146460 171633: contig of 25173 bp in length
* 171633 171732: gap of unknown length
* 171733 194849: contig of 23117 bp in length
* 194850 223128: gap of unknown length
* 223129 223228: contig of 28179 bp in length
* 223229 259005: gap of unknown length
* 259006 259105: contig of 35777 bp in length
* 259106 303189: gap of unknown length
* 303189 303189: contig of 44084 bp in length.
*
* Location/Qualifiers
* 1. 303189
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="3"
* /clone="RP11-594C8"
* 3466..3565
* /estimated_length=unknown
* 7535..7634

```

gap	/estimated_length=unknown
gap	13995. .14094
gap	/estimated_length=unknown
gap	22001. .22100
gap	/estimated_length=unknown
gap	30723. .30822
gap	/estimated_length=unknown
gap	42635. .42734
gap	/estimated_length=unknown
gap	54976. .55075
gap	/estimated_length=unknown
gap	67040. .67139
gap	/estimated_length=unknown
gap	79748. .79847
gap	/estimated_length=unknown
gap	91972. .92071
gap	/estimated_length=unknown
gap	108786. .108885
gap	/estimated_length=unknown
gap	13001. .130110
gap	/estimated_length=unknown
gap	146360. .146458
gap	/estimated_length=unknown
gap	171633. .171732
gap	/estimated_length=unknown
gap	194850. .194949
gap	/estimated_length=unknown
gap	223129. .223228
gap	/estimated_length=unknown
gap	259006. .259105
gap	/estimated_length=unknown

	Query Match	14.7%	Score 58.6;	DB 12;	Length 303189;
	Best Local Similarity	57.5%;	Pred. No. 3	3e-08;	
	Matches 103;	Conservative 1;	Mismatches 75;	Indels 0;	Gaps 0;
Oy	110	GGCAGTAGGCGCTACCTTTTGTACTGTTTGATATATATATATATATACATACACACACAC 169			
Dd	204738	GGGCAGTAGCAAAATATCCAAACTATATCATTCATCATATATATATACACACACACAC 2046797			
Oy	170	ACACACACACACATATATTTTAAGTCGGGACATTAATGTACACACCCTATTCAAAGTAA 229			
Dd	204678	ACACACTCACACACATGAATGATATACATACATCATATATGTATATCACTATATATATAT 204619			
Oy	230	AAGTACTATCTCAGAAGTAAACATATACATCCGCTATTGTCTAATACCTGCAATTAAGC 288			
Dd	204618	TATACATATATATGAATTAATAATACATTAAGCTGAAAGTGAACAACCTGTAGAGCATCAGC 204560			
	RESULT 18				
	AC021491				
LOCUS	AC021491	175262 bp	DNA	linear	FRI 27-MAR-2002
DEFINITION	Homo sapiens chromosome 17, clone RP11-94C24, complete sequence.				
ACCESSION	AC021491				
VERSION	AC021491.16	GI:19747233			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 175262) Britten,B., Linton,L., Nusbaum,C. and Lander,E. Homo sapiens chromosome 17, clone RP11-94C24 Unpublished 2 (bases 1 to 175262) Britten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckertly,R., Bede,F., Bohuslavsky,I., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Chapel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					



Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGuirk, A., McKernan, K., McPherson, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Leary, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

## TITLE

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome

## REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margrie, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Milnova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome

## REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margrie, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Milnova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

Submitted (27-MAR-2002) Whitehead Institute/MIT Center for Genome

## JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

All repeats were identified using RepeatMasker:  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR

## FEATURES

source

Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L5368  
 Center clone name: 94\_C\_24  
 -----  
 Only the first 175.3 kilobases of this clone are being submitted.  
 The remainder overlaps accession number AC004707 [WICGR project  
 L326].

FEATURES	Location/Qualifiers
repeat_region	1..175262
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="17"
repeat_region	/map="17"
repeat_region	/clone="RP11-94C24"
repeat_region	/clone_lib="RPC1-11 Human Male BAC"
repeat_region	511..612
repeat_region	/rpt_family="CAGAGA)n"
repeat_region	921..950
repeat_region	/rpt_family="TTTTTC)n"
repeat_region	966..989
repeat_region	/rpt_family="(TA)n"
repeat_region	complement(11770..2004)
repeat_region	/rpt_family="MER102b"
repeat_region	complement(2005..2293)
repeat_region	/rpt_family="AluYo"
repeat_region	complement(2294..2392)
repeat_region	/rpt_family="MER102b"
repeat_region	complement(3017..3217)
repeat_region	/rpt_family="MIR"
repeat_region	3204..3280
repeat_region	/rpt_family="MIR"
repeat_region	3915..4060
repeat_region	/rpt_family="MER5A"
repeat_region	complement(4068..4249)
repeat_region	/rpt_family="MIR"
repeat_region	complement(7603..7696)
repeat_region	/rpt_family="L2"
repeat_region	8974..9095
repeat_region	/rpt_family="C-rich"
repeat_region	10187..10360
repeat_region	/rpt_family="MIR"
repeat_region	10401..10487
repeat_region	/rpt_family="L1MA7"
repeat_region	10488..10767
repeat_region	/rpt_family="AluXs"
repeat_region	10768..11186
repeat_region	/rpt_family="L1MA7"
repeat_region	11187..11461
repeat_region	/rpt_family="AluXs"
repeat_region	11465..11491
repeat_region	/rpt_family="(GAAA)n"
repeat_region	11492..12020
repeat_region	/rpt_family="L1MA7"
repeat_region	12021..12047
repeat_region	/rpt_family="AT-rich"
repeat_region	12050..12270
repeat_region	/rpt_family="MIR"
repeat_region	complement(12304..12404)
repeat_region	/rpt_family="MIR"
repeat_region	complement(14791..14995)
repeat_region	/rpt_family="MIR"
repeat_region	19566..19602
repeat_region	/rpt_family="(CA)n"
repeat_region	19732..19807
repeat_region	/rpt_family="MIR"
repeat_region	complement(22107..22298)
repeat_region	/rpt_family="MIR"
repeat_region	22772..22901
repeat_region	/rpt_family="MIR"



REFERENCE 2 (bases 1 to 149879)

Quality coverage: 6.5 in Q20 base; agalove-1p



\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1214: contig of 1214 bp in length
1215 1314: gap of 100 bp
1315 2781: contig of 1467 bp in length
2782 2881: gap of 100 bp
2882 4986: contig of 2105 bp in length
4987 5087: gap of 100 bp
5087 7848: contig of 2761 bp in length
7848 7948: gap of 100 bp
7948 10045: contig of 2098 bp in length
10046 66062: contig of 55917 bp in length
66063 66163: gap of 100 bp
66163 70153: contig of 3390 bp in length
70153 70252: gap of 100 bp
70252 76303: contig of 6050 bp in length
76303 76402: gap of 100 bp
76402 80928: contig of 4526 bp in length
80928 81028: gap of 100 bp
81028 86671: contig of 5643 bp in length
86671 86772: gap of 100 bp
86772 94792: contig of 8021 bp in length
94792 94892: gap of 100 bp
94892 118128: contig of 23236 bp in length
118128 118228: gap of 100 bp
118228 147108: contig of 28880 bp in length
147108 147209: gap of 100 bp
147209 149879: contig of 2671 bp in length.
  
```

## FEATURES

```

source
1..149879
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="8"
  /map="8"
  /clone="RP11-262F21"
  /clone_lib="RPC1-11 Human Male BAC"
  1..1214
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left"
  gap
  1215..1314
    /estimated_length=100
  misc_feature
  1315..2781
    /note="assembly_fragment"
    2782..2881
      /estimated_length=100
  gap
  2882..4986
    /note="assembly_fragment"
    4987..5086
      /estimated_length=100
  misc_feature
  5087..7847
    /note="assembly_fragment"
    7848..7947
      /estimated_length=100
  gap
  7948..10045
    /note="assembly_fragment"
    10046..10145
      /estimated_length=100
  misc_feature
  10146..66062
    /note="assembly_fragment"
    66063..66162
      /estimated_length=100
  gap
  66163..70152
    /note="assembly_fragment"
    70153..70252
  
```

```

misc_feature
70253..76302
  /note="assembly_fragment"
  76303..76402
    /estimated_length=100
  misc_feature
  76403..80928
    /note="assembly_fragment"
    80929..81028
      /estimated_length=100
  gap
  81029..86671
    /note="assembly_fragment"
    86672..86771
      /estimated_length=100
  misc_feature
  86772..94792
    /note="assembly_fragment"
    94793..94892
      /estimated_length=100
  misc_feature
  94893..118128
    /note="assembly_fragment"
    118129..118228
      /estimated_length=100
  misc_feature
  118229..147108
    /note="assembly_fragment"
    147109..147208
      /estimated_length=100
  gap
  147209..149879
    /note="assembly_fragment"
    clone_end:77
    vector_side:right"
  
```

## ORIGIN

Query Match 14.2%; Score 56.6; DB 12; Length 149879;  
 Best Local Similarity 64.3%; Pred. No. 1.4e-07;  
 Matches 83; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

```

QY 122 CTACCTTTTGTACTGTTGATATATATATATATACATACACACACACACACACACAC
DB 54215 CACTATATTTTCATTTCTCATATATATATACACACACACACACACACACAC
QY 182 AATATATTTTAACTCGGACATATATATATATATATATATATATATATATATAT
DB 54155 ATATATTTTCAATATATATATATATATATATATATATATATATATATATATAT
QY 242 AGAAGTAA 250
DB 54095 AGAAGATA 54087
  
```

```

RESULT 20
AL929213/c 169897 bp DNA linear VRT 28-NOV-2002
LOCUS Zebrafish DNA sequence from clone CH211-144A19, complete sequence.
DEFINITION AL929213
ACCESSION AL929213
VERSION AL929213.5 GI:25955864
KEYWORDS HMG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
REFERENCE 1 (bases 1 to 169897)
AUTHORS Dunn,M.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
COMMENT On Nov 29, 2002 this sequence version replaced gi:25045550.
  
```

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zface@sanger.ac.uk  
 -----







```

* 20527 22655: contig of 2129 bp in length
* 22656 22755: gap of 100 bp
* 22756 26278: contig of 3523 bp in length
* 26278 26379: gap of 100 bp
* 26379 28018: contig of 1640 bp in length
* 28018 32278: contig of 100 bp
* 32278 32379: contig of 4160 bp in length
* 32379 37065: contig of 100 bp
* 37065 37165: gap of 100 bp
* 37165 47028: contig of 9864 bp in length
* 47028 47129: gap of 100 bp
* 47129 58786: contig of 11658 bp in length
* 58786 58887: gap of 100 bp
* 58887 72062: contig of 13175 bp in length
* 72062 72161: gap of 100 bp
* 72161 88242: contig of 16081 bp in length
* 88242 88343: gap of 100 bp
* 88343 107620: contig of 19278 bp in length
* 107620 107721: gap of 100 bp
* 107721 130141: contig of 22421 bp in length
* 130141 130241: gap of 100 bp
* 130241 163955: contig of 33714 bp in length
* 163955 164056: gap of 100 bp
* 164056 211076: contig of 47021 bp in length.

```

## FEATURES

## SOURCE

```

1. 211076
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="8"
  /map="8"
  /clone="RP11-304J22"
  /clone_id="RP11-11 Human Male BAC"
  1. 1266
    /note="assembly_fragment"
    1267. 1366
      /estimated_length=100
      /note="assembly_fragment"
    1367. 2507
      /estimated_length=100
      /note="assembly_fragment"
    2508. 2607
      /estimated_length=100
      /note="assembly_fragment"
    2608. 3625
      /note="assembly_fragment"
    3630. 3729
      /estimated_length=100
      /note="assembly_fragment"
    3730. 5506
      /estimated_length=100
      /note="assembly_fragment"
    5507. 5606
      /estimated_length=100
      /note="assembly_fragment"
    5607. 7205
      /estimated_length=100
      /note="assembly_fragment"
    7206. 7305
      /estimated_length=100
      /note="assembly_fragment"
    7306. 8960
      /note="assembly_fragment"
    8961. 9060
      /estimated_length=100
      /note="assembly_fragment"
    9061. 11532
      /estimated_length=100
      /note="assembly_fragment"
    11533. 11632
      /estimated_length=100
      /note="assembly_fragment"
    11633. 12962
      /note="assembly_fragment"
    12963. 13062
      /estimated_length=100
      /note="assembly_fragment"
    13063. 15359
      /note="assembly_fragment"
    15360. 15459
      /estimated_length=100
      /note="assembly_fragment"
    15460. 16817
      /note="assembly_fragment"
    16818. 16917

```

```

misc_feature 16918..18580
               /note="assembly_fragment"
gap 18581..18680
               /estimated_length=100
misc_feature 18681..20426
               /note="assembly_fragment"
gap 20427..20526
               /estimated_length=100
misc_feature 20527..22655
               /note="assembly_fragment"
gap 22656..22755
               /estimated_length=100
misc_feature 22756..26278
               /note="assembly_fragment"
gap 26279..26378
               /estimated_length=100
misc_feature 26379..28018
               /note="assembly_fragment"
clone_end:SP6
vector_end:right"
28019..28118
               /estimated_length=100
28119..32378
               /note="assembly_fragment"
clone_end:T7
vector_end:left"
32379..32378
               /estimated_length=100
32379..37064
               /note="assembly_fragment"
37065..37164
               /estimated_length=100
37165..47028
               /note="assembly_fragment"
47029..47128
               /estimated_length=100
47129..58786
               /estimated_length=100

Query Match 14.2%, Score 56.6; DB 12; Length 211076;
Best Local Similarity 64.3%; Pred. No. 1.5e-07;
Matches 83; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

Cy 122 CTACCTTTTCTACTGTTGATATATATATATACATACACACACACACACAC 181
    |||||
Db 99967 CAATATATTCGCAATTTCTCATATATATACACACACACACACACACACAC 100026

Cy 182 AATATTTTAACTCGGGGCAATTAATGTACACACCTATTCAAAGTAAACTACTATCTC 241
    |||||
Db 100027 AATATTTTCAATATCATATATATATTTGGCCAGACACTGAAGGAAATGAAGA 100086

Cy 242 AGAAGTAA 250
    |||||
Db 100087 AGAAGATA 100095

RESULT 22
AC098683 112727 bp DNA linear ROD 09-MAR-2004
LOCUS AC098683
DEFINITION Mus musculus strain 129/Sv clone ctf-128f23 map 5, complete
sequence.
AC098683
VERSION AC098683.18 GI:45267891
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 112727)
Do,T., Drabek,B. and Roe,B.A.
Mus musculus BAC Clone ctf-128f23
JOURNAL Unpublished

```



REFERENCE 2 (bases 1 to 112727)  
AUTHORS Do,T., Drabek,B. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (30-OCT-2001) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 112727)  
AUTHORS Do,T., Drabek,B. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUL-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 4 (bases 1 to 112727)  
AUTHORS Do,T., Drabek,B. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUL-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 5 (bases 1 to 112727)  
AUTHORS Do,T., Drabek,B. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (21-FEB-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 6 (bases 1 to 112727)  
AUTHORS Do,T., Drabek,B. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Mar 9, 2004 this sequence version replaced gi:42734074.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKMOR  
-----  
FEATURES  
source  
1. 112727  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="129/Sv"  
/db\_xref="taxon:10090"  
/map="5"  
/clone="cct-128f23"  
/clone\_lib="c1boc7 mouse BAC library"  
-----  
ORIGIN  
Query Match 14.1%; Score 56.2; DB 6; Length 112727;  
Best Local Similarity 70.9%; Pred.No. 1.9e-07;  
Matches 73; Conservative 1; Mismatches 29; Indels 0; Gaps 0;  
-----  
QY 90 TCTGGAAGTAGATGAGGCGGCAAGCCCTTTGTACTGTTGATATAT 149  
Db 82837 TCTTTAAAAAAATCCAAACAAAGCCGCTTACGACAGACCTTGTCTCAATAT 82896  
QY 150 ATATACATACACACACACACACACACACACACATATATTTAA 192  
Db 82897 ATATACACACACACACACACACACACACACATATATTTCA 82939  
-----  
RESULT 23  
AC060796/c 118646 bp DNA linear PRI 19-FEB-2002  
LOCUS Homo sapiens chromosome 17, clone CTD-2245L18, complete sequence.  
ACCESSION AC060796  
VERSION AC060796.8 GI:18702411  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 118646)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone CTD-2245L18  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 118646)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Baetsen,V., Bada,F.,  
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
Campiano,A., Casale,A., Choepel,Y., Collinge,M., Collins,S.,  
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
Galand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamm,L., Karakas,A.,  
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Riley,R., Rogov,P., Rothman,D.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,  
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,  
Teafaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
JOURNAL Submitted (20-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 118646)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
Anderson,S., Barna,N., Baetsen,V., Boguslavsky,L., Boukhalter,B.,  
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karakas,A., Kells,J.C., Larocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
McClennan,C., MacDonald,P., Major,J., Margulis,N., Mathew,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,  
Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
Oliver,J., Peterson,K., Phunthang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,  
Straus,N., Subramanian,A., Talmas,J., Teafaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
JOURNAL Submitted (11-SEP-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 4 (bases 1 to 118646)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
Anderson,S., Barna,N., Baetsen,V., Boguslavsky,L., Boukhalter,B.,  
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karakas,A., Kells,J.C., Larocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., McClennan,C.,  
MacDonald,P., Major,J., Margulis,N., Mathew,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrum,J., Menus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunthang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,  
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,











Puzzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Riley, B., Reilly, M., Ren, Y., Reutter, M., Richards, S., Riggs, F.,  
Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sadders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
Shetty, J., Sivartsbeyn, A., Sibson, I., Slitter, C. D., Snajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,  
Valdes, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

\*\*\*\*\*

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 236644)  
AUTHORS Morley, K.C.  
TITLE Direct Submission  
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 236644)

\*\*\*\*\*

REFERENCE Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 10, 2003 this sequence version replaced gi:24918703.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

\*\*\*\*\*

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GSZF  
Center clone name: CH230-33NG  
Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 228522 bases at least Q40  
Consensus quality: 229930 bases at least Q30  
Consensus quality: 231245 bases at least Q20  
Estimated insert size: 243865; sum-of-contigs estimation  
Quality coverage: 12x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\*

-- NOTE: Estimated insert size may differ from sequence length  
-- (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
-- NOTE: This is a "working draft" sequence. It currently  
-- consists of 6 contigs. The true order of the pieces  
-- is not known and their order in this sequence record is  
-- arbitrary. Gaps between the contigs are represented as  
-- runs of N, but the exact sizes of the gaps are unknown.  
-- This record will be updated with the finished sequence  
-- as soon as it is available and the accession number will  
-- be preserved.

\*\*\*\*\*

1 229587: contig of 229587 bp in length  
\* 229588 229687: gap of unknown length  
\* 229688 231130: contig of 1443 bp in length  
\* 231131 231230: gap of unknown length  
\* 231231 232260: contig of 1030 bp in length







[illegible][illegible]



**AUTHORS** Wilson, R.K.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (25-JUN-2005) Washington University School of Medicine, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

**COMMENT** On May 27, 2005 this sequence version replaced gi:61741042.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.wustl.edu  
 ----- Summary Statistics  
 Center project name: C\_AD-656P08

#### NOTICE:

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Chimpanzee Chromosome Y Mapping Project (Jennifer F. Hughes, Tatyana Pynlikova, Helen Skalecky, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

**source**  
 1..179557  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /chromosome="Y"  
 /clone="CH251-656P8"  
 15385..15406  
 /note="Sequence derived from one plasmid subclone."  
 21563..21644  
 /note="Sequence derived from one plasmid subclone."  
 24731..24794  
 /note="Sequence derived from one plasmid subclone."  
 63005..64876  
 /note="Sequence derived from PCR product of project DNA"  
 69710..69742  
 /note="Sequence derived from one plasmid subclone."  
 73358..73434  
 /note="Sequence derived from one plasmid subclone."  
 73595..73820  
 /note="Sequence derived from one plasmid subclone."  
 74479..74480  
 /note="Bacterial transposon insertion in clone excised here"  
 86018..86029  
 /note="Sequence derived from one plasmid subclone."  
 86018..86029

#### ORIGIN

Query Match 14.0%; Score 56; DB 5; Length 179557;  
 Best Local Similarity 65.6%; Pred.No. 2.3e-07;  
 Matches 80; Conservative 1; Mismatches 41; Indels 0; Gaps 0;  
 138 TTGATATATATATATACACACACACACACACACACACACACATTTTAACTCGG 197  
 Db 159499 TTCTATATATATATATATACACACACACACACACACACACACACAAATTTGTG 159558

Oy 198 GACATAATGTACACACACCTTCAAGTAAAACTACTCTCAGAGTAAAGACATTA 257  
 Db 159559 GTTATATATCAAAATAGTCAGTTCAAACTTATTAAGCTTATGAGAAAAACCAAGACACA 159618  
 Oy 258 TC 259  
 Db 159619 CC 159620

#### RESULT 30

BS000534/c 224014 bp DNA linear PRI 17-MAR-2004  
 LOCUS  
 DEFINITION Pan troglodytes chromosome Y clone:PTB-092H12, complete sequence.  
 ACCESSION BS000534  
 VERSION BS000534.1 GI:45504134  
 HTG.

#### KEYWORDS

HTG.  
 Pan troglodytes (chimpanzee)  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Pan.

**REFERENCE**  
 1 The Chimpanzee Chromosome Y Sequencing Consortium.  
 DNA sequence of chimpanzee chromosome Y and its evolutionary implications

**REFERENCE**  
 2 (bases 1 to 224014)  
 Hattori, M., Toyoda, A., Noguchi, H., Taylor, T.D., Kuroki, Y., Fujiiyama, A. and Sakaki, Y.

**REFERENCE**  
 Direct Submission  
 Submitted (15-MAR-2004) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://ngp.gsc.riken.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

#### JOURNAL

**TITLE** The Chimpanzee Chromosome Y Sequencing Consortium.  
**AUTHORS** Hattori, M., Toyoda, A., Noguchi, H., Taylor, T.D., Kuroki, Y., Fujiiyama, A. and Sakaki, Y.  
**JOURNAL** Submitted (15-MAR-2004) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://ngp.gsc.riken.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

#### FEATURES

**source**  
 1..224014  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /chromosome="Y"  
 /clone="PTB-092H12"

#### ORIGIN

Query Match 14.0%; Score 56; DB 5; Length 224014;  
 Best Local Similarity 65.6%; Pred.No. 2.4e-07;  
 Matches 80; Conservative 1; Mismatches 41; Indels 0; Gaps 0;  
 138 TTGATATATATATATATACACACACACACACACACACACACACATTTTAACTCGG 197  
 Db 223417 TTCTATATATATATATATACACACACACACACACACACACACACAAATTTGTG 223358  
 Oy 198 GACATAATGTACACACCTTCAAGTAAAACTACTCTCAGAGTAAAGACATTA 257  
 Db 223357 GTTATATATCAAAATAGTCAGTTCAAACTTATTAAGCTTATGAGAAAAACCAAGACACA 223298  
 Oy 258 TC 259  
 Db 223297 CC 223296

#### RESULT 31

AC146193 232304 bp DNA linear PRI 19-MAY-2004  
 LOCUS  
 DEFINITION Pan troglodytes BAC clone CH251-483N24 from Y, complete sequence.  
 ACCESSION AC146193  
 VERSION AC146193.3 GI:47131439  
 HTG.

#### KEYWORDS

HTG.  
 Pan troglodytes (chimpanzee)  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homiidae; Pan.



REFERENCE 1 (bases 1 to 232304)  
 AUTHORS Shahid,S., Haglund,K. and Bielicki,L.  
 TITLE The sequence of Pan troglodytes BAC clone CH251-483N24  
 JOURNAL Unpublished (2001)  
 REFERENCE 2 (bases 1 to 232304)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 232304)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-APR-2004) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 232304)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAY-2004) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 232304)  
 AUTHORS Wilson,R.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-MAY-2004) Washington University School of Medicine, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 COMMENT On May 12, 2004 this sequence version replaced gi:4659556.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: MUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.edu  
 ----- Summary Statistics  
 Center project name: C\_AB0483N24  
 -----  
 NOTICE:  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
 -----  
 MAPPING INFORMATION:  
 The position of this clone was established as part of a collaboration between the Chimpanzee Chromosome Y Mapping Project (Jennifer P. Hughes, Tatyana Pyntikova, Helen Skalecky, Steve Rosen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.  
 -----  
 SOURCE INFORMATION:  
 The CHORI-251 Chimpanzee BAC library has been constructed at the Children's Hospital Oakland Research Institute, BACPAC Resources, by Dr. Baoji Zhu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate:6-6-80). The clone and detailed information can be obtained from Pieter de Jong and co-workers at http://www.bacpac.chori.org.  
 -----  
 NEIGHBORING SEQUENCE INFORMATION:  
 This sequence is the entire insert of the clone. This clone is overlapped by AC147341.  
 -----  
 FEATURES  
 SOURCE  
 1. 232304  
 Location/Qualifiers  
 /organism="Pan troglodytes"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9598"  
 /chromosome="Y"  
 /map="Y"  
 /clone="CH251-483N24"

Query Match 14.0%; Score 56; DB 5; Length 232304;  
 Best Local Similarity 65.6%; Pred. No. 2.4e-07;  
 Matches 80; Conservative 1; Mismatches 41; Indels 0; Gaps 0;  
 ORIGIN  
 Query 138 TTGATATATATATATATACATACACACACACACACACACATATCTGAGTAAGCATTA 197  
 Db 93777 TTCATATATATATATATATACACACACACACACACACACACACATTTGG 93836  
 Qy 198 GACATTAATGACACACCTTCAAGTAATAAATCTATCTGAGTAAGCATTA 257  
 Db 93837 GTTATATCAATTAAGTCACTTCAACTTATTAAGTATGAGAAAACCAAGACACA 93896  
 Qy 258 TC 259  
 Db 93897 CC 93898  
 RESULT 32  
 AC025227 60810 bp DNA linear PRI 09-MAY-2001  
 LOCUS AC025227  
 DEFINITION Homo sapiens BAC clone RP11-109G18 from Y, complete sequence.  
 ACCESSION AC025227.6 GI:11120937  
 VERSION AC025227.6  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 60810)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 REFERENCE 2 (bases 1 to 60810)  
 AUTHORS Swearingen,S., Abbot,A. and Doeber,A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-109G18  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 60810)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 60810)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 60810)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission



JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Nov 8, 2000 this sequence version replaced gi:931947.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@genome.wustl.edu](mailto:sapiens@genome.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0109G18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:  
The RPL1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatem, M., Calanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RPL1-95B23, 200 bp overlap; the clone sequenced to the right is RPL1-477B5, 200 bp overlap. Actual start of this clone is at base position 195 of RPL1-95B23; actual end is at base position 54795 of RPL1-477B5.

FEATURES  
Location/Qualifiers

1..60810

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="Y"

/map="Y"

/clone="RPL1-109G18"

/clone\_1lb="RPL1-11"

1..85

/rpt\_family="Alu"

repeat\_region

265..560

/rpt\_family="Alu"

repeat\_region

2076..2380

/rpt\_family="L1"

repeat\_region

2450..2738

/rpt\_family="L1"

repeat\_region

2745..2803

/rpt\_family="Alu"

repeat\_region

3684..3957

/rpt\_family="Alu"

repeat\_region

4467..4758

/rpt\_family="Alu"

repeat\_region

5026..5165

/rpt\_family="MALR"

repeat\_region 5167..5331  
/rpt\_family="Alu"  
repeat\_region 5332..5494  
/rpt\_family="L1"  
repeat\_region 5495..5624  
/rpt\_family="Alu"  
repeat\_region 5625..5773  
/rpt\_family="MALR"  
repeat\_region 5777..6089  
/rpt\_family="ERV1"  
repeat\_region 6187..6275  
/rpt\_family="ERV1"  
repeat\_region 6488..6734  
/rpt\_family="Alu"  
repeat\_region 6735..7039  
/rpt\_family="ERV1"  
repeat\_region 7084..7256  
/rpt\_family="Alu"  
repeat\_region 7792..7840  
/rpt\_family="ERV1"  
repeat\_region 11645..11947  
/rpt\_family="Alu"  
repeat\_region 12684..12930  
/rpt\_family="Alu"  
repeat\_region 13098..13393  
/rpt\_family="Alu"  
repeat\_region 15135..15248  
/rpt\_family="Alu"  
repeat\_region 15249..15704  
/rpt\_family="ERV1"  
repeat\_region 15810..16105  
/rpt\_family="Alu"  
repeat\_region 16168..17185  
/rpt\_family="ERV1"  
repeat\_region 18582..18676  
/rpt\_family="ERV1"  
repeat\_region 19702..19790  
/rpt\_family="L1"  
repeat\_region 19822..19889  
/rpt\_family="L1"  
repeat\_region 20026..20093  
/rpt\_family="L1"  
repeat\_region 20116..20168  
/rpt\_family="L1"  
repeat\_region 20456..21092  
/rpt\_family="L1"  
repeat\_region 21147..21588  
/rpt\_family="L1"  
repeat\_region 21629..21913  
/rpt\_family="L1"  
repeat\_region 21944..22564  
/rpt\_family="Alu"  
repeat\_region 22369..22679  
/rpt\_family="L1"  
repeat\_region 22696..23103  
/rpt\_family="L1"  
repeat\_region 23362..23512  
/rpt\_family="L1"  
repeat\_region 23851..24119  
/rpt\_family="Alu"  
repeat\_region 24347..24425  
/rpt\_family="ERV"  
repeat\_region 24466..24973  
/rpt\_family="ERV"  
repeat\_region 28368..29123  
/rpt\_family="ERV1"  
repeat\_region 29132..29333  
/rpt\_family="ERV1"  
repeat\_region 30293..30562  
/rpt\_family="ERV"  
repeat\_region 30570..30920  
/rpt\_family="MALR"  
repeat\_region 31587..32003







REFERENCE 3 (bases 1 to 212207)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 212207)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 212207)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Oct 11, 2001 this sequence version replaced gi:1578784.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0502M01  
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,B., Tatehno,M., Catalano,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the right is RP11-219G14. Actual start of this clone is at base position 1 of RP11-502M1; actual end is at base position 212207 of RP11-502M1.

The sequence of AC013337 has been incorporated into AC093853.

FEATURES  
 source  
 Location/Qualifiers  
 1..212207  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-502M1"  
 /clone\_id="RPCI-11"  
 repeat\_region  
 239..408

repeat\_region  
 410..536  
 /rpt\_family="L2"  
 /rpt\_family="L2"  
 891..1138  
 /rpt\_family="L2"  
 repeat\_region  
 1658..1694  
 /rpt\_family="AT\_rich"  
 1695..1768  
 /rpt\_family="TA)n"  
 2079..2362  
 /rpt\_family="TA)n"  
 2377..2616  
 /rpt\_family="TA)n"  
 4159..4293  
 /rpt\_family="MIR"  
 4906..5042  
 /rpt\_family="MER1\_type"  
 5740..6146  
 /rpt\_family="L2"  
 6853..6981  
 /rpt\_family="TA)n"  
 7233..7524  
 /rpt\_family="Alu"  
 8043..8532  
 /rpt\_family="TA)n"  
 8652..8724  
 /rpt\_family="ERV1"  
 8656..8722  
 /rpt\_family="CA)n"  
 8770..8946  
 /rpt\_family="L1"  
 9683..9712  
 /rpt\_family="AT\_rich"  
 9813..9837  
 /rpt\_family="AT\_rich"  
 12004..12524  
 /rpt\_family="MER1\_type"  
 12505..12558  
 /rpt\_family="AT\_rich"  
 12594..12615  
 /rpt\_family="MER1\_type"  
 12615..12645  
 /rpt\_family="(T)n"  
 12616..12926  
 /rpt\_family="Alu"  
 12927..13003  
 /rpt\_family="MER1\_type"  
 13262..13631  
 /rpt\_family="MAIR"  
 13632..15201  
 /rpt\_family="MAIR"  
 15202..15564  
 /rpt\_family="MAIR"  
 15700..15746  
 /rpt\_family="L2"  
 15858..16368  
 /rpt\_family="BRV1"  
 16461..17334  
 /rpt\_family="BRVX"  
 17348..17734  
 /rpt\_family="BRVX"  
 17901..17928  
 /rpt\_family="(TG)n"  
 19062..19133  
 /rpt\_family="AT\_rich"  
 19172..19193  
 /rpt\_family="AT\_rich"  
 19928..20049  
 /rpt\_family="L2"  
 20068..20477  
 /rpt\_family="L1"  
 20832..21061  
 /rpt\_family="L1"



Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dethorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedetich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, K.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durkin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabali, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hognus, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovac, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Y., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, M., Louisedg, H., Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinec, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Moore, S., Morgan, M., Moorishi, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newsum, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguni, M., Okumura, G., Oragunye, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Piktums, E., Pu, L., Qulles, M., Ren, Y., Rivers, M., Rojaja, A., Rojibokhan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitrali, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Szatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Taylor, B., Thomas, N., Thomas, S., Umanu, K., Vasquez, L., Vera, V., Villallon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zylator, S.L., Zelnick, G., and Gibbs, R.

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 176108)
AUTHORS	Worley, K.C.

JOURNAL	Submitted (19-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 176108)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (29-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 176108)
AUTHORS	Worley, K.C.

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Apr 30, 2002 this sequence version replaced gi:20336718.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

**ANNOTATION OF FEATURES:**

STs were identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.



```
repeat_region      /rpt_family="AT_rich"
                   14756. .14963
```

REFERENCE AUTHORS	TITLE
1 (bases 1 to 184182)	DOE Joint Genome Institute, Alamos National Laboratory.
2 (bases 1 to 184182)	Direct Submission
DOE Joint Genome Institute.	Unpublished



[illegible]

```

FEATURES
SOURCE
    Location/Qualifiers
        1..668
            /organism="Meleagris galllopavo"
            /mol_type="genomic DNA"
            /db_xref="taxon:9103"
            1..668
                /standard_name="TC1458"

ORIGIN
Query Match      13.8%; Score 55.2; DB 7; Length 668;
Best Local Similarity 58.9%; Pred. No. 1.9e-07;
Matches 93; Conservative 1; Mismatches 64; Indels 0; Gaps 0;

QY 129 TTGTGACTGTTGATATATATATATATACATACACACACACACACACACATATATT 188
    |||
Db 217 TTTCTTATCTGTGATATATATACATATACACACACACACACACACACACACATATA 158
    |||

QY 189 TTAACTCGGGGCATAAATGTACACCACCTATTCAAGTAATAAATACTATCTCGAAGTA 248
    |||
Db 157 TATATGCATATGTATCATATTCATCAGGAATTTATATACATCAAGAAAATCGCAAAGTC 98
    |||

QY 249 AAGACATPATCTCTCTATTGTTCATTAACTGACAGTTAA 286
    |||
Db 97 TAGTGAAAAATCTGGACACCTGATCATGAAGAAGTACAA 60
    |||

RESULT 38
LOCUS AL593618 668 bp DNA linear STS 16-JUL-2001
DEFINITION Meleagris galllopavo DNA; STS TC1458, sequence tagged site.
ACCESSION AL593618
VERSION ALS93618.1 GI:14716736
KEYWORDS STS.
SOURCE Meleagris galllopavo (turkey)
ORGANISM Meleagris galllopavo
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.

REFERENCE
AUTHORS Morrice,D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 668)
AUTHORS Morrice,D.
TITLE Direct Submision
COMMENT Submitted (16-JUL-2001) Morrice D., Roslin Institute, Roslin,
Midlothian, EH25 9PS, UK david.morrice@bsrc.ac.uk
Vector pBLUESCRIPT SK+
R. Site 1 EcoRI
R. Site 2 EcoRI
Contact: David Morrice
Dept. Genomics and Bioinformatics
Roslin Institute, Roslin
Midlothian, EH25 9PS, UK
Tel: +44 (0) 131 527 4200
Fax: +44 (0) 131 440 0434
Email: david.morrice@bsrc.ac.uk
Seq primer 17.

FEATURES
SOURCE
    Location/Qualifiers
        1..668
            /organism="Meleagris galllopavo"
            /mol_type="genomic DNA"
            /db_xref="taxon:9103"
            1..668
                /standard_name="TC1458"

ORIGIN
Query Match      13.8%; Score 55.2; DB 7; Length 668;
Best Local Similarity 58.9%; Pred. No. 1.9e-07;
Matches 93; Conservative 1; Mismatches 64; Indels 0; Gaps 0;

STS
1..668
/standard_name="TC1458"


```



DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	REFERENCE	AUTHORS
0y	AC099909/c	Mus musculus clone RP23-13B5, low-pass sequence sampling.	AC099909	AC099909.1	GI:17047275	HTG; HTGS PHASE0	Mus musculus (house mouse)	1 (bases 1 to 61255)	Birren, B., Linton, L., Nusbaum, C. and Lander, E.	Unpublished	2 (bases 1 to 61255)	Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Baetjen, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopio, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kelle, C., Lacroque, K., Lamares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J., Menzies, L., Milova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Recta, R., Riebeck, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stefanovic, N., Struss, N., Subramanian, A., Talamas, J., Testay, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, W.
0y	TTTACCTCGGACATTAATGTACACCACTATTCAAAGTAAAGTAACTATCTCTGAAAGTA	189										
Db	TATATGCATATGTTACATATTCATGAGAAATTTATATACATCAAGAAATACGACAAAGTC	512										
0y	249 AAGACATATCTCTGATTCATTCATTAATACAGCTTA	286										
Db	572 TAGGAAAAATCTGGACACCTGATCATGAAACTGACAA	609										
COMMENT	<p>NOTE: This record contains 76 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for</p>											

- \* identifying clones that may be gene-rich and allows
- \* overlap relationships among clones to be deduced.
- \* However, it should not be assumed that this clone
- \* will be sequenced to completion. In the event that
- \* the record is updated, the accession number will
- \* be preserved.

1	648:	conf: of 648 bp	in length
649	748:	gap of 100 bp	
749	1470:	conf: of 722 bp	in length
1471	1570:	gap of 100 bp	
1571	2275:	conf: of 705 bp	in length
2276	2375:	gap of 100 bp	
2376	3059:	conf: of 684 bp	in length
3060	3159:	gap of 100 bp	
3160	3850:	conf: of 691 bp	in length
3851	3950:	gap of 100 bp	
3951	4659:	conf: of 709 bp	in length
4660	4759:	gap of 100 bp	
4760	5440:	conf: of 681 bp	in length
5441	5540:	gap of 100 bp	
5541	6241:	conf: of 721 bp	in length
6241	6314:	gap of 100 bp	
6362	7082:	conf: of 701 bp	in length
7063	7162:	gap of 100 bp	
7163	7873:	conf: of 711 bp	in length
7874	7973:	gap of 100 bp	
7974	8683:	conf: of 710 bp	in length
8684	8783:	gap of 100 bp	
8784	9486:	conf: of 713 bp	in length
9497	9596:	gap of 100 bp	
9597	10300:	conf: of 704 bp	in length
10301	10400:	gap of 100 bp	
10401	11119:	conf: of 715 bp	in length
11120	11219:	gap of 100 bp	
11220	11919:	conf: of 700 bp	in length
11920	12019:	gap of 100 bp	
12020	12749:	conf: of 730 bp	in length
12750	12849:	gap of 100 bp	
12850	13555:	conf: of 706 bp	in length
13556	13655:	gap of 100 bp	
13656	14377:	conf: of 722 bp	in length
14378	14477:	gap of 100 bp	
14478	15189:	conf: of 712 bp	in length
15190	15289:	gap of 100 bp	
15290	16000:	conf: of 711 bp	in length
16001	16100:	gap of 100 bp	
16101	16808:	conf: of 708 bp	in length
16809	16908:	gap of 100 bp	
16909	17622:	conf: of 714 bp	in length
17623	17722:	gap of 100 bp	
17723	18404:	conf: of 682 bp	in length
18405	18504:	gap of 100 bp	
18505	19210:	conf: of 706 bp	in length
19211	19310:	gap of 100 bp	
19311	20020:	conf: of 710 bp	in length
20021	20120:	gap of 100 bp	
20121	20839:	conf: of 719 bp	in length
20840	20939:	gap of 100 bp	
20940	21631:	conf: of 698 bp	in length
21638	21731:	gap of 100 bp	
21738	22447:	conf: of 710 bp	in length
22448	22547:	gap of 100 bp	
22548	23265:	conf: of 718 bp	in length
23266	23365:	gap of 100 bp	
23366	24070:	conf: of 705 bp	in length
24071	24170:	gap of 100 bp	
24171	24860:	conf: of 690 bp	in length
24861	24960:	gap of 100 bp	
24961	25671:	conf: of 711 bp	in length
25672	25771:	gap of 100 bp	
25772	26481:	conf: of 710 bp	in length
26482	26581:	gap of 100 bp	
26582	27257:	conf: of 676 bp	in length



```

* 27258 27357: gap of 100 bp
* 27358 28067: contig of 720 bp in length
* 28068 28167: gap of 100 bp
* 28168 28876: contig of 709 bp in length
* 28877 28976: gap of 100 bp
* 28977 29681: contig of 705 bp in length
* 29682 29781: gap of 100 bp
* 29782 30488: contig of 707 bp in length
* 30489 30588: gap of 100 bp
* 30589 31289: contig of 701 bp in length
* 31290 31390: gap of 100 bp
* 31390 32112: contig of 723 bp in length
* 32113 32212: gap of 100 bp
* 32213 32911: contig of 699 bp in length
* 32912 33011: gap of 100 bp
* 33012 33724: contig of 713 bp in length
* 33725 33824: gap of 100 bp
* 33825 34521: contig of 697 bp in length
* 34522 34621: gap of 100 bp
* 34622 35321: contig of 700 bp in length
* 35322 35421: gap of 100 bp
* 35422 36128: contig of 707 bp in length
* 36129 36228: gap of 100 bp
* 36229 36945: contig of 717 bp in length
* 36946 37045: gap of 100 bp
* 37046 37757: contig of 712 bp in length
* 37758 37857: gap of 100 bp
* 37858 38559: contig of 702 bp in length
* 38560 38659: gap of 100 bp
* 38660 39351: contig of 692 bp in length
* 39352 39451: gap of 100 bp
* 39452 40159: contig of 708 bp in length
* 40160 40259: gap of 100 bp
* 40260 40956: contig of 697 bp in length
* 40957 41056: gap of 100 bp
* 41057 41776: contig of 720 bp in length
* 41777 41876: gap of 100 bp
* 41877 42592: contig of 716 bp in length
* 42593 42692: gap of 100 bp
* 42693 43397: contig of 705 bp in length
* 43398 43497: gap of 100 bp
* 43498 44195: contig of 698 bp in length
* 44196 44295: gap of 100 bp
* 44296 45030: contig of 735 bp in length
* 45031 45130: gap of 100 bp
* 45131 45841: contig of 711 bp in length
* 45842 45941: gap of 100 bp
* 45942 46660: contig of 719 bp in length
* 46661 46760: gap of 100 bp
* 46761 47473: contig of 713 bp in length
* 47474 47573: gap of 100 bp
* 47574 48287: contig of 714 bp in length
* 48288 48387: gap of 100 bp
* 48388 49102: contig of 715 bp in length
* 49103 49202: gap of 100 bp
* 49203 49908: contig of 706 bp in length
* 49909 50008: gap of 100 bp
* 50009 50736: contig of 728 bp in length
* 50737 50836: gap of 100 bp
* 50837 51527: contig of 691 bp in length
* 51528 51627: gap of 100 bp
* 51628 52327: contig of 700 bp in length
* 52328 52427: gap of 100 bp
* 52428 53145: contig of 718 bp in length
* 53146 53245: gap of 100 bp
* 53246 53974: contig of 729 bp in length
* 53975 54074: gap of 100 bp
* 54075 54789: contig of 715 bp in length
* 54790 54889: gap of 100 bp
* 54890 55586: contig of 697 bp in length

```

Query Match  
Best Local Similarity

13.8% Score 55.2; DB 12; Length 61255;  
66.1% Pred. No. 3.7e-07;

```

Matches 78; Conservative 1; Mismatches 39; Indels 0; Gaps 0;
Qy 119 GACCTACCTTTTGTACTGTTGATATATATATATACACACACACACACA 178
Db 9818 GGTCTACGATACATGTAAATATACACACACACACACACACACACACA 9759
Qy 179 CACATATTTTAACTCGGACATTAATGTCACACACCTATTCAAGTAAACTACT 236
Db 9758 CACACAAACATTCATCGCACATTAATATGCTACATATATGCAACATACATCACT 9701

RESULT 40
AL590437
LOCUS AL590437 151358 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-394E12, 6 unordered pieces.
ACCESSION AL590437
VERSION AL590437.4 GI:13620473
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1
REFERENCE
AUTHORS McLay, K.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Apr 12, 2001 this sequence version replaced gi:13619301.

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba394E12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 4% of reads
Dye-terminator Big Dye; 95% of reads
Consensus quality: 149170 bases at least Q40
Consensus quality: 149671 bases at least Q30
Consensus quality: 150169 bases at least Q20
Insert size: 148776; 6.0% error; agarose-fp
Quality coverage: 7.08x in Q20 bases; sum-of-contigs Quality
coverage: 7.18x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 22354: contig of 22354 bp in length
* 22355 22454: gap of 100 bp
* 22455 27105: contig of 4651 bp in length
* 27106 27205: gap of 100 bp
* 27206 47237: contig of 20032 bp in length
* 47238 47337: gap of 100 bp
* 47338 64870: contig of 17533 bp in length
* 64871 64970: gap of 100 bp
* 64971 74038: contig of 9068 bp in length
* 74039 74138: gap of 100 bp
* 74139 151358: contig of 77220 bp in length.
Location/Qualifiers
1. .151358
/organism="Homo sapiens"
/mol_type="genomic DNA"
FEATURES
source

```







repeat_region	3579..3637	/rpt_family="Alu"
repeat_region	3663..3848	/rpt_family="L2"
repeat_region	4553..4654	/rpt_family="L2"
repeat_region	4651..4755	/rpt_family="L2"
repeat_region	5631..8264	/rpt_family="L1"
repeat_region	9547..9852	/rpt_family="MER1_type"
repeat_region	9990..10054	/rpt_family="MER1_type"
repeat_region	10252..10282	/rpt_family="AT_rich"
repeat_region	10279..10478	/rpt_family="MER1_type"
repeat_region	10615..10640	/rpt_family="AT_rich"
repeat_region	10671..10963	/rpt_family="Alu"
repeat_region	10957..10990	/rpt_family="(A)n"
repeat_region	12173..12277	/rpt_family="Tc2"
repeat_region	13689..13997	/rpt_family="L1"
repeat_region	14211..14412	/rpt_family="L2"
repeat_region	14411..14729	/rpt_family="L2"
repeat_region	14777..15003	/rpt_family="MR"
repeat_region	15039..15403	/rpt_family="MER2_type"
repeat_region	15400..15510	/rpt_family="(PATA)G)n"
repeat_region	15506..15629	/rpt_family="L2"
repeat_region	15630..15830	/rpt_family="MR"
repeat_region	15866..15979	/rpt_family="L2"
repeat_region	16071..16491	/rpt_family="MeLR"
repeat_region	17261..17507	/rpt_family="Mariner"
repeat_region	18707..19811	/rpt_family="L1"
repeat_region	19686..19717	/rpt_family="AT_rich"
repeat_region	19812..20111	/rpt_family="Alu"
repeat_region	20112..20956	/rpt_family="L1"
repeat_region	20965..21424	/rpt_family="L1"
repeat_region	21463..21522	/rpt_family="AT_rich"
repeat_region	21567..21816	/rpt_family="L1"
repeat_region	21890..22003	/rpt_family="MER1_type"
repeat_region	22037..22135	/rpt_family="MER1_type"
repeat_region	22576..22615	/rpt_family="(T)n"
repeat_region	22650..22668	/rpt_family="(TA)n"
repeat_region	22686..22715	/rpt_family="(CA)n"
repeat_region	29023..29196	

repeat_region	/rpt_family="MER1_type"	29127. .29151	/rpt_family="AT_rich"	31344. .31706	/note="similar to EST BF512614 (NID:g11597793)"
misc_feature		31641. .33029			
repeat_region	/rpt_family="L2"	32184. .33382	/rpt_family="L2"	32403. .32543	
repeat_region	/rpt_family="L2"	32403. .32543	/rpt_family="CR1"	32562. .32781	
repeat_region	/rpt_family="MIR"	32803. .32949	/rpt_family="CR1"	32931. .33395	
misc_feature	/note="similar to EST AW236561 (NID:g6568950) xm47g11.x1"	34086. .34219			
repeat_region	/rpt_family="MIR"	34117. .34473	/note="similar to EST AA889061 (NID:g3015940) am34h02.s1"	34965. .37833	
repeat_region	/rpt_family="L1"	37854. .38167			
repeat_region	/rpt_family="A1u"	38704. .38737			
Query Match	13.8%;	Score 55.2;	DB 5;	Length 190080;	
Best Local Similarity	70.6%;	Pred. No. 4,4e-07;			
Matches 77;	Conservative 1;	Mismatches 29;	Indels 0;	Gaps 0;	

[illegible]

RESULT	42
Locus	AC136742/c
Definition	Mus musculus chromosome 5, clone RP23-145H8, complete sequence.
Accession	AC136742
Version	AC136742.15
Keywords	HTG.
Source	Mus musculus (house mouse)
Organism	Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

2 (bases 1 to 197131)

Mus musculus chromosome 5, clone RE23-145H8

Unpublished

2 (bases 1 to 197131)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bassleten, V., Bloom, T., Boguslavsky, L., Bonkshgiller, B., Camarata, J., Chang, J., Chazaro, S., Choepel, Y., Collamore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gird, S., Graham, L., Grand-Pierre, N., Hafez, N., Haggoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karakas, A., Kellis, C., Lander, C., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menues, L., Mhova, T., Miensva, V., Murphy, T., Newell, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnuppach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testafaye, S., Theodore, J., Topham, K.,







Allen, C., Allen, H., Alsbrooke, S., Amin, A., Anguiano, D.,  
Altena, E., Altschuler, V., Aoyagi, A., Ayodell, M., Bacc, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnesstead, M., Barnamed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Bulah, C., Burch, C., Burckell, K., Calderon, B.,  
Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., De Souza, L.,  
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denison, S., Detramo, C., Ding, Y., Dinh, H., Diya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Dublin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P., Foster, P.,  
Frazer, C. M., Gabriel, A., Gante, R., Garcia, A., Garner, T., Garza, M.,  
Geisler, E., Gertel, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
Hollins, L., Jacob, L., Jiang, R., Johnson, B., Johnson, R., Jolivet, A.,  
Karapathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.

Kowls, C., Kafri, C.L., Lebow, H., Levran, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loru, L., Loru, L., Louieged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawlawi, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nwakolameli, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K., Pasternack, S., Paul, H., Perez, A., Perez, L., Plambeck, C., Plopper, F., Polindexter, A., Popovic, D., Primm, E., Pu, L., Rizzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scheer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Stemple, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, P., Williams, G., Willson, R., Wlaczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 235887)

Worley, K.C.

Direct Submission

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235887)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24819555. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GHY

Center clone name: CH230-133C19

----- Summary Statistics

Assembly program: Atlas 3.0.1

Consensus quality: 21393 bases at least Q40

Consensus quality: 21747 bases at least Q30

Consensus quality: 220158 bases at least Q20

Estimated insert size: 228571; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently



\* consist of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preverred.  
\* 1 235887: contig of 235887 bp in length.



















GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

On nucleic - nucleic search, using sw model

Run on: July 17, 2006, 16:49:59 ; Search time 3964 Seconds  
(without alignment)  
703.557 Million cell updates/sec

Title: SEQ1-33670G  
Perfect score: 399.2  
Sequence: 1 aagaatgcctctctctcata.....taactcagatcagaacgag 400

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues  
Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 150 summaries

Database :

N Geneseq. 8: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004as: \*  
13: Geneseqn2004bs: \*  
14: Geneseqn2005s: \*  
15: Geneseqn2006s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	397.6	99.6	83400	12	ADP079906		Adp079906 Human. RAD
	2	397.6	99.6	193363	14	AED89392		Aed89392 Human bre
	3	60.8	15.2	265703	14	AEE04792		Aee04792 Cancer-as
	4	56	14.0	463	4	AAI92919		Aai92919 Human pol
	5	53.6	14.0	463	9	ACH24278		Ach24278 Human adu
	6	53.6	13.4	141463	11	ACN43862		Acn43862 Human adu
	7	52.4	13.1	3455	5	AAH41011		Aah41011 Human gen
	8	51.6	12.9	327	3	ADP57118		Adp57118 Human met
	9	51.2	12.8	49914	13	ABD33262		Abd33262 Murine ca
	10	51	12.8	3462	5	AAH41010		Aah41010 Human met
	11	51	12.8	3464	5	AAH41006		Aah41006 Human met
	12	51	12.8	3467	5	AAH41005		Aah41005 Human met
	13	51	12.8	3467	5	AAH41009		Aah41009 Human met
	14	51	12.8	3469	5	AAH41007		Aah41007 Human met
	15	51	12.8	3470	5	AAH41008		Aah41008 Human met
	16	51	12.8	3473	5	AAH41004		Aah41004 Human met
	17	51	12.8	24000	3	AAH88551		Aah88551 Human dua
	18	51	12.8	28854	8	ABX95685		Abx95685 Human gen

c	19	50.8	12.7	92726	9	ADA02927		Ada02927 Mouse pik
c	20	50.8	12.7	92726	10	ADB72665		Adb72665 Mouse pik
c	21	50.8	12.7	92726	10	ADC85407		Acd85407 Mouse pik
c	22	50.8	12.7	92726	12	ADM74522		Adm74522 Murine pik
c	23	50.2	12.6	101169	12	ADQ97584		Adq97584 Mouse can
c	24	50	12.5	5289	4	AAK67814		Aak67814 Human imm
c	25	50	12.5	61876	11	ACN44732		Acn44732 Mouse gen
c	26	50	12.5	68960	14	ADZ13128		Adz13128 Murine ca
c	27	50	12.5	110000	11	ACN43998		Acn43998 Murine ca
c	28	49.8	12.5	151127	10	ADH13788		Adh13788 Prostate
c	29	49.8	12.5	2534	3	AAC69141		Aac69141 Human ABC
c	30	49.8	12.5	17970	6	ABQ67033		Abq67033 Human ang
c	31	49.8	12.5	28829	13	ABD33193		Abd33193 Murine ca
c	32	49.8	12.5	159865	12	ADQ97056		Adq97056 Human can
c	33	49.8	12.5	188056	15	AEP19127		Aep19127 Human NIP
c	34	49.4	12.4	590	13	ADQ57023		Adq57023 Novel can
c	35	49.4	12.4	4858	6	ABK33962		Abk33962 Human DNA
c	36	49.4	12.4	4858	8	ADA20413		Ada20413 Prostate
c	37	49.4	12.4	4858	8	ADA84220		Ada84220 Human ren
c	38	49.4	12.4	110000	10	ADE11169		Adel1169
c	39	49.4	12.4	154394	12	ADQ17784		Adq17784 Human sof
c	40	49.4	12.4	154394	14	ADX06954		Adx06954 Cyclin-de
c	41	49.4	12.4	154394	14	ADZ80338		Adz80338 Human nuc
c	42	49.4	12.4	154394	15	AEP74999		Aep74999 Human pol
c	43	49.2	12.3	13427	6	ABL33926		Abi33926 Human imm
c	44	49	12.3	109565	13	ABD33086		Abd33086 Murine ca
c	45	49	12.3	138627	12	ADQ97183		Adq97183 Human can
c	46	48.8	12.2	10138	6	ABK28343		Abk28343 DNA trans
c	47	48.8	12.2	40633	11	ACN43848		Acn43848 Mouse gen
c	48	48.8	12.2	110000	13	ABD32966		Abd32966
c	49	48.8	12.2	219352	13	ABD33098		Abd33098 Murine ca
c	50	48.8	12.2	247509	10	ADL13684		Adl13684 Osteoarth
c	51	48.8	12.2	247509	12	ADQ59419		Adq59419 Human can
c	52	48.6	12.2	247654	14	ADZ13712		Adz13712 Murine ca
c	53	48.6	12.2	350	4	AAI80148		Aai80148 Human pol
c	54	48.6	12.2	2316	14	ADY17693		Ady17693 DNA encod
c	55	48.6	12.2	110000	6	ABQ74964		Abq74964
c	56	48.6	12.2	156843	11	ACN44786		Acn44786 Human gen
c	57	48.4	12.1	3205	13	ADR07286		Adr07286 Full leng
c	58	48.4	12.1	3205	14	AED22221		Aed22221 Human DNA
c	59	48.4	12.1	18401	11	ADP75183		Adp75183 Human ADA
c	60	48.4	12.1	19236	6	ABN80227		Abn80227 Human che
c	61	48.4	12.1	20099	10	ABZ22599		Abz22599 Human F13
c	62	48.4	12.1	54108	6	ABK23782		Abk23782 Human hig
c	63	48.4	12.1	57273	6	ABK22784		Abk22784 Human hig
c	64	48.4	12.1	66933	5	ABA82625		Aba82625 Human HBM
c	65	48.4	12.1	66933	8	ACA45366		Acc45366 Human HBM
c	66	48.4	12.1	66933	10	ADB88066		Adb88066 HBM-relat
c	67	48.4	12.1	66933	10	ADB82435		Adb82435 Human DNA
c	68	48.4	12.1	66933	13	ADR16929		Adr16929 BAC clone
c	69	48.4	12.1	66933	13	ADR47580		Adr47580 BAC clone
c	70	48.4	12.1	66933	14	AEB69309		Aeb69309 Human H1g
c	71	48.4	12.1	72049	5	ABA82623		Abas82623 Human HBM
c	72	48.4	12.1	72049	8	ACC45364		Acc45364 Human HBM
c	73	48.4	12.1	72049	10	ADB88064		Adb88064 HBM-relat
c	74	48.4	12.1	72049	10	ADB82433		Adb82433 Human DNA
c	75	48.4	12.1	72049	13	ADR16927		Adr16927 BAC clone
c	76	48.4	12.1	72049	13	ADR47578		Adr47578 BAC clone
c	77	48.4	12.1	72049	14	AEB69307		Aeb69307 Human H1g
c	78	48.4	12.1	110000	13	AEE04936		Aee04936
c	79	48.4	12.1	137046	14	ABD29936		Abd29936
c	80	48.4	12.1	210920	12	ADQ97123		Adq97123 Mouse can
c	81	48.2	12.1	1544	13	ADVA1681		Adva1681 Rat cardl
c	82	48.2	12.1	1544	14	AEC11878		Aec11878 Rat surto
c	83	48.2	12.1	1544	14	AEC12529		Aeci12529 Rat surto
c	84	48.2	12.1	118502	11	ACN44112		Acn44112 Mouse gen
c	85	48.2	12.1	197526	13	ABD33389		Abd33389 Human can
c	86	48	12.0	5952	4	AA641403		Aa641403 Human gen
c	87	48	12.0	32217	4	AA641738		Aa641738 Genomic s
c	88	48	12.0	91609	12	ADQ59473		Adq59473 Human can
c	89	48	12.0	91609	14	ADZ13834		Adz13834 Murine can
c	90	47.8	12.0	378	2	AAV87853		Aav87853 EST clone
c	91	47.8	12.0	1329	2	AAV33199		Aav33199 Secreted



C	92	47.8	12.0	3035	5	ADG20677	ADG20677 Human ABC	XX	Cytosolic; Gene Therapy; breast cancer; RAD21; human; chromosome 8q24;
C	93	47.8	12.0	7378	5	ADG20614	ADG20614 Human ABC	KW	gene; single nucleotide polymorphism; SNP; ds.
C	94	47.8	12.0	7379	6	ABL58158	ABL58158 Human ABC	KW	
C	95	47.8	12.0	31477	11	ACN44096	ACN44096 Mouse gen	XX	Homo sapiens.
C	96	47.8	12.0	32359	11	ACN44720	ACN44720 Mouse gen	OS	
C	97	47.8	12.0	38564	13	ADT77132	ADT77132 Type II d	XX	
C	98	47.8	12.0	55008	4	AAK80495	AAK80495 Human imm	XX	
C	99	47.8	12.0	108316	10	ADG87336	ADG87336 Human GPC	FT	Key
C	100	47.8	12.0	164702	8	ACF62730	ACF62730 Cancer ba	FT	variation
C	101	47.8	12.0	164702	8	ADB20845	ADB20845 MRP1 base	FT	location/Qualifiers
C	102	47.8	12.0	164702	10	ADB87934	ADB87934 Human UGT	FT	replace(56,C)
C	103	47.8	12.0	164702	10	ADB96917	ADB96917 Human MDR	FT	replace(7324,A)
C	104	47.8	12.0	164702	10	ADB92108	ADB92108 Human MDR	FT	replace(7324,A)
C	105	47.8	12.0	326014	6	ABK89296	ABK89296 Human gen	FT	replace(7324,A)
C	106	47.8	12.0	326014	12	ADQ94981	ADQ94981 Human kin	FT	replace(7324,A)
C	107	47.6	11.9	426	14	ADM83621	ADM83621 MAP3K9 ma	FT	replace(9231,T)
C	108	47.6	11.9	3657	10	ABZ22598	ABZ22598 Human FL3	FT	replace(9231,T)
C	109	47.6	11.9	47804	10	ADC86176	ADC86176 Human GPC	FT	replace(9231,T)
C	110	47.6	11.9	87311	11	ACN44792	ACN44792 Mouse gen	FT	replace(10490,A)
C	111	47.6	11.9	110000	12	ADQ97266	ADQ97266 Human can	FT	replace(10490,A)
C	112	47.6	11.9	220895	6	ABK84798	ABK84798 Human CDN	FT	replace(10490,A)
C	113	47.6	11.9	220895	13	ADRS2737	ADRS2737 Drug ther	FT	replace(11867,G)
C	114	47.6	11.9	349881	10	ADC86642	ADC86642 Human GPC	FT	replace(11867,G)
C	115	47.6	11.9	349938	10	ADC87621	ADC87621 Human GPC	FT	replace(11867,G)
C	116	47.4	11.9	389	2	AAV90181	AAV90181 EST clone	FT	replace(13308,A)
C	117	47.4	11.9	4361	4	ABL04348	ABL04348 Drosophi1	FT	replace(13308,A)
C	118	47.4	11.9	13217	4	AAAS35088	AAAS35088 DNA #38 e	FT	replace(13308,A)
C	119	47.4	11.9	13217	4	AAAS31477	AAAS31477 Human DNA	FT	replace(13308,A)
C	120	47.4	11.9	13217	4	AAI63997	AAI63997 Human pol	FT	replace(13308,A)
C	121	47.4	11.9	13217	4	ABA06818	ABA06818 Human gen	FT	replace(13308,A)
C	122	47.4	11.9	13217	5	ABA18587	ABA18587 Human ner	FT	replace(13308,A)
C	123	47.4	11.9	13217	5	AA829749	AA829749 Human end	FT	replace(13308,A)
C	124	47.4	11.9	13217	6	ABQ66801	ABQ66801 Human pol	FT	replace(13308,A)
C	125	47.4	11.9	13217	6	ABV84155	ABV84155 Human pol	FT	replace(13308,A)
C	126	47.4	11.9	13217	10	ADCI1088	ADCI1088 Human DNA	FT	replace(13308,A)
C	127	47.4	11.9	13217	10	ADC46530	ADC46530 Human neo	FT	replace(13308,A)
C	128	47.4	11.9	13217	12	ADM24712	ADM24712 Human PRO	FT	replace(13308,A)
C	129	47.4	11.9	13220	4	AAAS35086	AAAS35086 DNA #36 e	FT	replace(13308,A)
C	130	47.4	11.9	13220	4	AAAS31475	AAAS31475 Human DNA	FT	replace(13308,A)
C	131	47.4	11.9	13220	4	AAI63995	AAI63995 Human pol	FT	replace(13308,A)
C	132	47.4	11.9	13220	4	ABA06816	ABA06816 Human gen	FT	replace(13308,A)
C	133	47.4	11.9	13220	5	ABA18585	ABA18585 Human ner	FT	replace(13308,A)
C	134	47.4	11.9	13220	5	AA829747	AA829747 Human end	FT	replace(13308,A)
C	135	47.4	11.9	13220	6	ABQ66799	ABQ66799 Human pol	FT	replace(13308,A)
C	136	47.4	11.9	13220	6	ABV84153	ABV84153 Human pol	FT	replace(13308,A)
C	137	47.4	11.9	13220	10	ADCI1086	ADCI1086 Human DNA	FT	replace(13308,A)
C	138	47.4	11.9	13220	10	ADC46528	ADC46528 Human neo	FT	replace(13308,A)
C	139	47.4	11.9	13220	12	ADM24710	ADM24710 Human PRO	FT	replace(13308,A)
C	140	47.4	11.9	73308	6	ABL66966	ABL66966 Lung canc	FT	replace(13308,A)
C	141	47.4	11.9	183999	4	AAF92831	AAF92831 Human ABC	FT	replace(13308,A)
C	142	47.4	11.9	183999	13	ADU76470	ADU76470 Human ABC	FT	replace(13308,A)
C	143	47.2	11.8	1068	6	ABE65590	ABE65590 Mouse gen	FT	replace(13308,A)
C	144	47.2	11.8	6652	6	ABQ66988	ABQ66988 Human ang	FT	replace(13308,A)
C	145	47.2	11.8	17967	6	ABL33015	ABL33015 Human imm	FT	replace(13308,A)
C	146	47.2	11.8	31477	11	ACN44096	ACN44096 Mouse gen	FT	replace(13308,A)
C	147	47.2	11.8	76341	12	ADM97422	ADM97422 Prostata	FT	replace(13308,A)
C	148	47.2	11.8	133632	11	ACN45054	ACN45054 Human gen	FT	replace(13308,A)
C	149	47.2	11.8	337022	12	ADQ59416	ADQ59416 Human can	FT	replace(13308,A)
C	150	47.2	11.8	499	4	AAI86291	AAI86291 Human pol	FT	replace(13308,A)

## ALIGNMENTS

RESULT 1	ADP07906	ADP07906 standard; DNA; 83400 BP.
XX	ADP07906;	
AC	26-AUG-2004	(first entry)
XX	Human RAD21 gene, SEQ ID 1.	
XX		
DE		



FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(33545,G)
FT		/*tag= w
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(33670,C)
FT		/*tag= x
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(33778,C)
FT		/*tag= y
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(34691,T)
FT		/*tag= z
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(36859,A)
FT		/*tag= aa
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(47651,G)
FT		/*tag= ab
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(48463,A)
FT		/*tag= ac
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(49058,C)
FT		/*tag= ad
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(50233,A)
FT		/*tag= ae
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(51228,G)
FT		/*tag= af
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(52315,G)
FT		/*tag= ag
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(53039,G)
FT		/*tag= ah
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(62940,T)
FT		/*tag= ai
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(64531,C)
FT		/*tag= aj
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(64989,G)
FT		/*tag= ak
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(65209,C)
FT		/*tag= al
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(65444,C)
FT		/*tag= am
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(70056,T)
FT		/*tag= an
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(70329,G)
FT		/*tag= ao
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(70629,G)
FT		/*tag= ap
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(71326,A)
FT		/*tag= aq
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(72563,A)
FT		/*tag= ar
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(73135,T)
FT		/*tag= as
FT	variation	/standard_name="Single Nucleotide Polymorphism"
FT		replace(73627,A)
FT		/*tag= at
FT	variation	/standard_name="Single Nucleotide Polymorphism"

Query Match	Best Local Similarity	99.6% Score	397.6; DB 12; Length 83400;
Matches	397; Conservative	2; Mismatches	1; Indels 0; Gaps 0;
QY 1	AAGAATGCTCTTCTCTATATATTTTACGAGTTCTTGAATAGCTTAATACCTGTGGCG	60	
DB 33596	AAGAATGCTCTTCTCTATATATTTTACGAGTTCTTGAATAGCTTAATACCTGTGGCG	33655	
QY 61	TACACATGAGCTGCTGATCTGCTTACCTCTGGAAGAGTATGACGGGCAAGGAGG	120	
DB 33666	TACACATGAGCTGCTGATCTGCTTACCTCTGGAAGAGTATGACGGGCAAGGAGG	33715	
QY 121	CCTACCTTTTGTACTGTTTGTATATATATATATACATACACACACACACACACA	180	
DB 33726	CCTACCTTTTGTACTGTTTGTATATATATATATACATACACACACACACACACA	33775	
QY 181	CATATATTTTAACTCGGACATTAATGTACACACACCTTCAAGTAAAACTATATCT	240	
DB 33776	CATATATTTTAACTCGGACATTAATGTACACACACCTTCAAGTAAAACTATATCT	33835	
QY 241	CAGAAGTAAAGCATTAATCTGCTATTTGTCACTTAATCTAGCAGTTAAGCTGCTAACCTTT	300	
DB 33836	CAGAAGTAAAGCATTAATCTGCTATTTGTCACTTAATCTAGCAGTTAAGCTGCTAACCTTT	33895	
QY 301	ATGCAAGTTCAGTTTCCCTACTATATGAAAACTGAAAGGTTGAATCTGCTGTTTACTTAA	360	















The invention relates to novel foetal urogenital sinus-derived expressed sequence tags. The invention more specifically comprises an isolated polynucleotide comprising a nucleotide sequence containing a urogenital sinus-derived expressed sequence tag (EST) comprising g932, ug993, ug101, ug102, ug106, ug120, ug254, ug329, ug307, ug308, ug311, ug317, ug320, g9334, ug335, ug353, ug354, ug357, ug440, ug441, ug482, ug484, ug485, ug491, ug493, ug494, ug503, ug505, ug506, ug818, ug8186 and ug8194. The urogenital sinus-derived expressed sequence tags have cytosinatic and anti-inflammatory activities. The genes containing urogenital sinus-derived ESTs encode oncological proteins that are useful for diagnosis of prostate disease and for monitoring the efficacy of a compound in clinical trials for the treatment of prostate disease. The ESTs can be used in assays to identify substances (e.g. triple helix forming oligonucleotides, antisense oligonucleotides, ribozymes or antibodies) useful for treating

Query Match	Score	DB	Length
12.8%	51.2	13	49914







Pt Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N,  
DR WPI, 2001-343602/36.  
XX  
XX Metalloprotease with aggrecanase activity for treating joint diseases  
PT especially osteoarthritis.  
PT  
Claim 11; Page 76-77; 85pp; Japanese.  
PS  
PS This invention relates to a metalloprotease with aggrecanase activity.  
CC The invention includes protein and DNA sequences of the metalloprotease,  
CC vectors containing the DNA, host cells transformed by the vectors, and  
CC antibodies directed against the metalloprotease. The antibodies, protein  
CC and DNA sequences can be used in the treatment and prevention of joint  
CC diseases, particularly osteoarthritis. The treatment may result in  
CC osteoplastic and antirheumatic activity. The present sequence represents a  
CC promoter related to the metalloprotease of the invention

SQ Sequence 3467 BP; 819 A; 895 C; 893 G; 860 T; 0 U; 0 Other;  
  
Query Match 12.8%; Score 51; DB 5; Length 3467;  
Best Local Similarity 82.6%; Pred. No. 0.0013;  
Matches 57; Conservative 1; Mismatches 11; Indels 0; Gaps 0

OY 121 CCTACCTTTTGTGACTGGTTGAATATATATATATATATACATTCACACACACACACACACA 180  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 2418 CATCATATATATATACATATCATATATATATATATATATACATTCACACACACACACACACA 2355  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
OY 181 CAVATATT 189  
||:||||| |  
DB 2358 CACATATAT 2350

RESULT 14  
 AAH41007/c  
 ID AAH41007 standard; DNA; 3469 BP.  
 XX  
 AC AAH41007;  
 XX  
 DT 23-AUG-2001 (first entry)  
 XX  
 DE Human metalloprotease related promoter SEQ ID 27.  
 XX  
 KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;  
 KW osteopathic; antiarthritic; promoter; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200134785-A1.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 10-NOV-2000; 2000WO-JP007917.  
 XX  
 PR 11-NOV-1999; 99JP-00321740.  
 PR 16-MAY-2000; 2000JP-00144020.  
 XX  
 PA (YAMA ) YAMANOUCHI PHARM CO LTD.  
 PA (KAZU-) KAZUSA DNA RES INST.  
 XX  
 PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;  
 DR WPI; 2001-343602/36.  
 XX  
 PT Metalloprotease with aggrecanase activity for treating joint diseases  
 XX especially osteoarthritis.  
 XX  
 PS Claim 11; Page 72-74; 85pp; Japanese.  
 XX  
 CC This invention relates to a metalloprotease with aggrecanase activity.  
 CC The invention includes protein and DNA sequences of the metalloprotease,  
 CC vectors containing the DNA, host cells transformed by the vectors, and







22-JAN-2001 (first entry)

Human dual-specificity phosphatase-1 (DSP-1) gene.

DSP-1; dual-specificity phosphatase-1; human; cell proliferation; cell differentiation; cell survival; cell cycle; dephosphorylation; signal transduction; MAP-kinase; cancer; graft versus host disease; allergy; autoimmune disease; metabolic disease; therapy; chromosome 17; ds.

Homo sapiens.

Key Location/Qualifiers

exon 42..109

intron /tag= a

intron /number= 1

intron 110..20823

exon /tag= b

exon /number= 1

exon 20824..21034

exon /tag= d

exon /number= 2a

exon /note= "Alternative, extended version of exon 2"

exon /tag= c

exon 20824..20911

intron /number= 2

intron /tag= e

intron /number= 2

intron 21035..22327

intron /tag= f

exon /number= 2a

exon 22328..23309

exon /tag= g

exon /number= 3

CDS 22420..23016

W0200053636-A2.

14-SEP-2000.

08-MAR-2000; 2000MO-US006154.

08-MAR-1999; 99US-0123255P.

(CEPT-) CEPTYR INC.

Lucho RM, Wei B;

WPI; 2000-579365/54.

P-PSDB; AAB19602.

New isolated polypeptide having the sequence of dual-specificity phosphatase-1 (DSP-1) is useful for treating a patient with a disorder associated with DSP-1 activity e.g. cancer and autoimmune diseases.

Example 1; Fig 3A-J; 74pp; English.

The present sequence is that of the human dual-specificity phosphatase-1 (DSP-1) gene on chromosome 17. The gene was identified in genomic sequences obtained from an expressed sequence tag database screened with a conserved motif (see AAB19604) of known DSPs. DSP-1 dephosphorylates both phosphoserine/threonine and phosphotyrosine residues in DSP-1 substrates such as activated mitogen-activated protein kinase (MAP-kinase). DSP-1 has sequence homology to other MAP-kinase phosphatases. It is expressed at high levels in the human heart, testis and liver, and at lower levels in other tissues. Methods are provided for recombinant production of DSP-1 polypeptides, and for using DSP-1 polypeptides, antibodies and polynucleotides to detect DSP-1 expression, to screen for agents that modulate DSP-1 activity e.g. within a combinatorial library, and for using such agents to modulate cell proliferation, cell differentiation or cell survival, through modulation of pattern of gene

expression, apoptosis or cell cycle. In particular, the cell displays contract inhibition of cell growth, anchorage-dependent growth or an altered intercellular adhesion property or is a cell present in a patient afflicted with a disorder associated with DSP-1 activity, such as cancer, graft-versus host disease, autoimmune disease, allergy, metabolic disease, abnormal cell growth, abnormal cell proliferation and abnormal cell cycle

Sequence 24000 BP; 6038 A; 5505 C; 5602 G; 6855 T; 0 U; 0 Other;

Query Match 12.8%; Score 51; DB 3; Length 24000;

Best Local Similarity 56.4%; Pred. No. 0.0023;

Matches 93; Conservative 1; Mismatches 71; Indels 0; Gaps 0;

94 GAAAGTAGTATGACGAGGCAAGTGGCTTCTTTTGTACTGTTGATATATATAT 153

19220 GAGTGAAAGAAAGCCAGCAAAAAAACCACATATATATATATATATATAT 19279

154 ACATACACACACACACACACACACACACACATATTTTAACGCGGACATTAATGACACC 213

19280 ATACACACACACACACACACACACACACACACACACACACCTTATATATATGTA 19339

214 ACCTATTCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 258

19340 GTATATGTGTACTATATATATATATATATATATATATATATATATATATAT 19384

RESULT 18

ID ABX95685/C

ABX95685 standard; DNA; 28854 BP.

ABX95685;

30-JUN-2003 (first entry)

Human gene encoding a protease of the metalloprotease family.

Human; ds; gene; chromosome 11; protease; metalloprotease; cancer; disintegrin; metalloprotease with thrombospondin motifs-1; cachectic; cytosolic; immunomodulator; SNP; single nucleotide polymorphism.

Homo sapiens.

Key Location/Qualifiers

variation replace(497,C)

variation /tag= q

variation /standard\_name= "Single nucleotide polymorphism"

variation replace(619,T)

variation /tag= r

variation /standard\_name= "Single nucleotide polymorphism"

variation replace(875,G)

variation /tag= s

CDS /standard\_name= "Single nucleotide polymorphism"

2001..26854

transcript= "protease"

transcript\_except= (pos:24100..24101,aa:Aaa)

2001..2957

exon /tag= b

intron /number= 1

intron 2958..14516

intron /tag= c

variation replace(3332,A)

variation /tag= t

variation /standard\_name= "Single nucleotide polymorphism"

variation replace(4208..4210,TG)

variation /tag= u

variation /standard\_name= "Single nucleotide polymorphism"

variation replace(4773,A)

variation /tag= v

variation /standard\_name= "Single nucleotide polymorphism"

variation replace(5009,G)



```

FT      /*tag= w
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(5597,G)
FT      /*tag= x
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(5664,C)
FT      /*tag= y
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(6323. .6324,TA)
FT      /*tag= z
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(6325,A)
FT      /*tag= aa
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(6564,G)
FT      /*tag= ab
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(10433. .10434,GGA)
FT      /*tag= ac
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(13184,G)
FT      /*tag= ad
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(13651,C)
FT      /*tag= ae
FT      /standard_name= "Single nucleotide polymorphism"
FT      14517. .14649
FT      /*tag= d
FT      /number= 2
FT      14650. .15114
FT      /*tag= e
FT      /number= 2
FT      15114. .15282
FT      /*tag= f
FT      /number= 3
FT      15283. .15524
FT      /*tag= g
FT      /number= 3
FT      replace(15471,G)
FT      /*tag= af
FT      /standard_name= "Single nucleotide polymorphism"
FT      15525. .15808
FT      /*tag= h
FT      /number= 4
FT      15809. .22292
FT      /*tag= i
FT      /number= 4
FT      replace(17255,T)
FT      /*tag= ag
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(17727,G)
FT      /*tag= ah
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(17833,G)
FT      /*tag= ai
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(18034. .18036,AA)
FT      /*tag= aj
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(18035,A)
FT      /*tag= ak
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(19087,A)
FT      /*tag= al
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(19371,G)
FT      /*tag= am
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(19478,C)
FT      /*tag= an
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(19803,A)
FT      /*tag= ao

```

```

FT      /standard_name= "Single nucleotide polymorphism"
FT      22293. .22470
FT      /*tag= j
FT      /number= 5
FT      replace(22448,C)
FT      /*tag= ar
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(22448,A)
FT      /*tag= ag
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(22448,G)
FT      /*tag= ap
FT      /standard_name= "Single nucleotide polymorphism"
FT      22471. .23952
FT      /*tag= k
FT      /number= 5
FT      replace(23568,T)
FT      /*tag= as
FT      /standard_name= "Single nucleotide polymorphism"
FT      replace(23820,A)
FT      /*tag= at
FT      /standard_name= "Single nucleotide polymorphism"
FT      23953. .24134
FT      /*tag= l
FT      /number= 6
FT      24135. .24240
FT      /*tag= m
FT      /number= 6
FT      24241. .24416
FT      /*tag= n
FT      /number= 7
FT      24417. .26079
FT      /*tag= o
FT      /number= 7
FT      replace(25713. .25715,CC)
FT      /*tag= au
FT      /standard_name= "Single nucleotide polymorphism"
FT      26080. .26851
FT      /*tag= p
FT      /number= 8
FT      exon
FT      US2002086400-A1.
FT      04-JUL-2002.
FT      21-DEC-2000; 2000US-00741151.
FT      06-DEC-2000; 2000US-0251398P.
FT      (ZHUS/) ZHU S.
FT      (GUEG/) GUEGLER K.
FT      (WEBS/) WEBSTER M.
FT      (DFRA/) DI FRANCESCO V.
FT      (BEAS/) BEASLEY E M.
FT      Zhu S, Guegler K, Webster M, Di Francesco V, Beasley EM;
FT      WPI: 2003-401333/38.
FT      P-PSDB; ABU09520.
FT      Novel human protease protein related to metalloprotease subfamily and
FT      nucleic acid molecule encoding the protein for diagnosing, treating
FT      disease or condition mediated by the protease protein e.g. cancer,
FT      cachexia.
FT      Claim 4; Fig 3; 64pp; English.
FT      The invention relates to an isolated human protease peptide (a member of
FT      the disintegrin/metalloprotease with thrombospondin motifs-1 family), an
FT      allelic variant or orthologue encoded by a nucleic acid molecule that
FT      hybridises under stringent conditions to the opposite strand of the cDNA
FT      and gene appearing as (ABX95684 and ABX95685) or a fragment of having 10
FT      contiguous amino acids. Also included are the encoding nucleic acids
CC

```



CC The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined  
CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasms, adenocarcinoma, and  
CC sarcomas. The present sequence represents a mouse gene of the invention.











PR	06-SEP-2000	2000US-0230437P	PR	17-NOV-2000	2000US-0249218P	PR	17-NOV-2004 (first entry)
PR	06-SEP-2000	2000US-0230438P	PR	17-NOV-2000	2000US-0249214P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0231242P	PR	17-NOV-2000	2000US-0249245P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0231243P	PR	17-NOV-2000	2000US-0249264P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0231244P	PR	17-NOV-2000	2000US-0249265P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0231413P	PR	17-NOV-2000	2000US-0249297P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0231414P	PR	17-NOV-2000	2000US-0249299P	PR	18-NOV-2004
PR	08-SEP-2000	2000US-0232060P	PR	17-NOV-2000	2000US-0249300P	PR	18-NOV-2004
PR	12-SEP-2000	2000US-0232081P	PR	01-DEC-2000	2000US-0250160P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0232397P	PR	05-DEC-2000	2000US-0250391P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0232398P	PR	05-DEC-2000	2000US-0251030P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0232399P	PR	05-DEC-2000	2000US-0251988P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0232400P	PR	06-DEC-2000	2000US-0256719P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0232401P	PR	06-DEC-2000	2000US-0251479P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0233063P	PR	08-DEC-2000	2000US-0251856P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0233064P	PR	08-DEC-2000	2000US-0251868P	PR	18-NOV-2004
PR	14-SEP-2000	2000US-0233065P	PR	08-DEC-2000	2000US-0251869P	PR	18-NOV-2004
PR	21-SEP-2000	2000US-0234023P	PR	08-DEC-2000	2000US-0251989P	PR	18-NOV-2004
PR	21-SEP-2000	2000US-0234274P	PR	08-DEC-2000	2000US-0251990P	PR	18-NOV-2004
PR	25-SEP-2000	2000US-0234977P	PR	11-DEC-2000	2000US-0254097P	PR	18-NOV-2004
PR	25-SEP-2000	2000US-0234988P	PR	05-JAN-2001	2001US-0259678P	PR	18-NOV-2004
PR	26-SEP-2000	2000US-0235484P	PR			PR	18-NOV-2004
PR	27-SEP-2000	2000US-0235834P	PR			PR	18-NOV-2004
PR	27-SEP-2000	2000US-0235836P	PR			PR	18-NOV-2004
PR	29-SEP-2000	2000US-0236337P	PR			PR	18-NOV-2004
PR	29-SEP-2000	2000US-0236367P	PR			PR	18-NOV-2004
PR	29-SEP-2000	2000US-0236368P	PR			PR	18-NOV-2004
PR	29-SEP-2000	2000US-0236369P	PR			PR	18-NOV-2004
PR	29-SEP-2000	2000US-0236370P	PR			PR	18-NOV-2004
PR	02-OCT-2000	2000US-0236802P	PR			PR	18-NOV-2004
PR	02-OCT-2000	2000US-0237037P	PR			PR	18-NOV-2004
PR	02-OCT-2000	2000US-0237038P	PR			PR	18-NOV-2004
PR	02-OCT-2000	2000US-0237039P	PR			PR	18-NOV-2004
PR	13-OCT-2000	2000US-0237040P	PR			PR	18-NOV-2004
PR	13-OCT-2000	2000US-0239355P	PR			PR	18-NOV-2004
PR	13-OCT-2000	2000US-0239357P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0240960P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241221P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241785P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241786P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241787P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241808P	PR			PR	18-NOV-2004
PR	20-OCT-2000	2000US-0241809P	PR			PR	18-NOV-2004
PR	01-NOV-2000	2000US-0244617P	PR			PR	18-NOV-2004
PR	08-NOV-2000	2000US-0					



















CC a mutation and is at least one of the mutations given in the  
CC specification. The gene has a mutation in an intron. The gene has a  
CC mutation in a promoter or regulatory region. The method is useful in  
CC diagnosing a patient with Cornelia de Lange Syndrome. The current  
CC sequence represents the genomic sequence of the NIPBL gene. This gene is  
CC located on human chromosome 5 in the region designated p13.1-13.3.  
XX  
SQ Sequence 188056 BP; 58089 A; 31468 C; 34382 G; 64117 T; 0 U; 0 Other;  
Query Match 12.5%; Score 49.8; DB 15; Length 188056;  
Best Local Similarity 50.2%; Pred. No. 0.0086;  
Matches 120; Conservative 1; Mismatches 118; Indels 0; Gaps 0;  
QY 20 ATATTTCGAGCTCTTGAATTAAGCTTAATACCTGTGGCTACACATGACATGTCAT 79  
DB 69597 ATTTAAACATTAATTTTGAATTAAGCTTAATACCTGTGGCTACACATGTCAT 69538  
QY 80 ACTGGTTACCTCTGGAAGTAGATGATGAGGCAAGTACCTTTTGTGCTGTT 139  
DB 69537 CACCTTTCCTAATGTTAATTTTGGTGGTGGGGAAGGCTTAACTAAGTACATAT 69478  
QY 140 TGATATATATATATATACATACACACACACACACACACACATATTTTAACTCGGA 199  
DB 69477 ATGTGTGTATATATATATATATACATACACACACACACATGATATATATACACAC 69418  
QY 200 CATTAATGTACACACACCTATTCAAAAGTAAACTACTCTCAGAAAGTAAAGACATAT 258  
DB 69417 ACAACACTGTTTGTGACAAAGAAAATTTAATATATATAGAAATTTAAGTAAAGAAAAT 69359  
RESULT 34  
AD057023/C  
ID AD057023 standard; DNA; 590 BP.  
AC  
AD057023;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Novel canine microarray-related DNA sequence SeqID8325.  
XX  
KM canine microarray; drug screening; toxicity assay;  
KM environmental pollutant; cellular response; gene expression profile;  
KM toxic response; liver necrosis; fatty liver disease;  
KM protein adduct formation; hepatitis; dog; ds.  
XX  
OS Canis familiaris.  
OS  
XX WO2004063324-A2.  
XX  
PN 29-JUL-2004.  
XX  
PB 05-MAY-2003; 2003WO-US013853.  
XX  
PE 03-MAY-2002; 2002US-0377240P.  
XX  
PR (GENE-) GENE LOGIC INC.  
PA (PRTZ ) PFIZER PROD INC.  
XX  
PI Digians JC, Porter M, Wei T;  
XX  
DR WPI; 2004-561890/54.  
XX  
XX New isolated nucleic acid molecule, useful for drug screening and  
XX toxicity assays or for assessing the impact, including toxicity, of a  
XX compound, pharmaceutical agent or environmental pollutant on a cell or  
XX living organism.  
XX  
PS Claim 1; SEQ ID NO 8325; 41bp; English.  
XX  
CC This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and

CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers  
CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver  
CC disease, protein adduct formation or hepatitis), those of the kidney,  
CC heart, brain or testes, or other pathologies associated with at least one  
CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention.  
XX  
SQ Sequence 590 BP; 201 A; 86 C; 84 G; 213 T; 0 U; 6 Other;  
Query Match 12.4%; Score 49.4; DB 13; Length 590;  
Best Local Similarity 56.0%; Pred. No. 0.0021;  
Matches 89; Conservative 2; Mismatches 68; Indels 0; Gaps 0;  
QY 68 GAGACTGTATCTGTTACTCTGGAAGTAGATGACGGGCACTGAGCTTACT 127  
DB 457 GATACCTGTTTATTTGATTCCTCTGAAATTAACAAAGCTTACTGATTTGTAC 398  
QY 128 TTTTGTACTGTTGAT 187  
DB 397 AT 338  
QY 188 TTTTACTCGGACATTAAT 226  
DB 337 GTCAAGAAAT 299  
RESULT 35  
ABK33962/C  
ID ABK33962 standard; DNA; 4858 BP.  
XX  
AC ABK33962;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human DNA for staging of Astrocytomas, complement, #23.  
XX  
KM Human; dog; astrocytoma; cytostatic; staging; cysteine methylation; CpG;  
KM bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;  
KM matrix assisted laser desorption/ionization mass spectrometry.  
XX  
OS Homo sapiens.  
OS  
XX WO200202808-A2.  
XX  
PN 10-JAN-2002.  
XX  
PD 02-JUL-2001; 2001WO-EP007538.  
XX  
PE 30-JUN-2000; 2000DE-01032529.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
PA (BPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;



WPI; 2002-171649/22.

Novel chemically modified genomic DNA sequences, useful in the characterization, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas or predisposition to astrocytomas.

Claim 1; SEQ ID NO 46; 37bp; English.

The invention relates to a nucleic acid comprising a sequence (I) of at least 18 bases in length of a segment of chemically pre-treated genomic DNA which has any one of the sequences of (ABX34032) or its complement. Also included are an oligonucleotide or peptide nucleic acid (or set thereof) of at least 9 nucleotides which hybridizes to (I), primers for (I), probes for detecting cytosine methylation or single-nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide nucleic acids for analysing diseases associated with the methylation states of the CpG dinucleotides of (I). The array is useful for determining genetic and/or epigenetic parameters, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas, or the predisposition to astrocytomas by analysing cytosine methylations, involves obtaining a biological sample containing genomic DNA, extracting the genomic DNA, converting cytosine bases which are unmethylated at the 5-position, in the genomic DNA sample, to uracil or another base which is dissimilar to cytosine in terms of hybridisation behaviour, by chemical treatment and amplifying chemically pre-treated genomic DNA fragments using the array and a polymerase, where the amplification carries a detectable label. The method further involves identifying methylation status of one or more cytosine positions, and analysing methylation status of the cytosine positions by reference to one or more data sets. The genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplification carries a fluorescent label or radionucleide. Optionally, the labels of the amplicifices are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplicifices or fragments of the amplicifices are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at <http://publ.int/publ/sequences>

Sequence 4858 BP; 995 A; 210 C; 1224 G; 2429 T; 0 U; 0 Other;

Query Match 12.4%; Score 49.4; DB 6; Length 4858;  
Best Local Similarity 52.8%; Pred. No. 0.0039;  
Matches 104; Conservative 1; Mismatches 92; Indels 0; Gaps 0;

```

QY 90 TCTGGAAGTAGTAGGAGGCAAGGAGGCTACCTTTTGTGACTGTTGATATATAT 149
DB 3711 TTTTAAACAATCCAAACAACAAAAAACACATCTCTACACACACACACACAC 3652
QY 150 ATATACATACACACACACACACACACACACACATATTTTAACTCGGACATTAATGTA 209
DB 3651 ACACACACACACACACACACACACACACACACACACAAATTAAGTATTAATACATA 3592
QY 210 CACCACTATTCAAGTAAAGTAACTATCTCAGAGTAAAGACATATCTGCTATTGT 269
DB 3591 AACCTATATATCCCACTATTAAACAAAAACAAAAAATTAATAAAAAATA 3532
QY 270 CATTAAGTACGAGTTAA 286
DB 3531 CCTTAACAAAAAATTA 3515

```

RESULT 36  
ADA20413/c

IDA20413 standard; DNA; 4858 BP.

ADA20413;

20-NOV-2003 (first entry)

Prostate tumour related genomic DNA complement sample #39.

Cytotoxic; gene therapy; genetic marker; epigenetic parameter;

classification; differentiation; diagnosis; prostate tumour;

prostate cancer; cytosine methylation; uracil;

single nucleotide polymorphism; SNP; prostate carcinoma; ss.

Homo sapiens.

W02002103042-A2.

27-DEC-2002.

14-JUN-2002; 2002MO-EP006605.

14-JUN-2001; 2001DE-01028508.

(EPIG-) EPIGENOMICS AG.

Distler J, Model F, Adorjan P;

WPI; 2003-167536/16.

Determining genetic and/or epigenetic parameters, useful for the classification, differentiation and/or diagnosis of prostate tumors or a predisposition to prostate cancer, comprises analyzing cytosine methylation.

Claim 28; Page 302-304; 376bp; English.

The invention relates to a method of determining genetic and/or epigenetic parameters for the classification, differentiation and/or diagnosis of prostate tumors or the predisposition to prostate cancer, by analysing cytosine methylation in a sample of genomic DNA. The method comprises chemically treating unmethylated cytosine bases at the 5-position to uracil or another base, which is dissimilar to cytosine in terms of hybridization behaviour, followed by amplifying at least one fragment of the chemically pre-treated genomic DNA using sets of primer oligonucleotides and a polymerase. The oligomers or probes derived from them are useful for detecting the methylation state of all CpG dinucleotides and/or single nucleotide polymorphisms (SNPs) in a chemically pre-treated genomic DNA. They are all useful for treating prostate carcinoma. This sequence represents a fragment of genomic DNA used in the method of the invention.

Sequence 4858 BP; 995 A; 210 C; 1224 G; 2429 T; 0 U; 0 Other;

Query Match 12.4%; Score 49.4; DB 8; Length 4858;  
Best Local Similarity 52.8%; Pred. No. 0.0039;  
Matches 104; Conservative 1; Mismatches 92; Indels 0; Gaps 0;

```

QY 90 TCTGGAAGTAGTAGGAGGCAAGGAGGCTACCTTTTGTGACTGTTGATATATAT 149
DB 3711 TTTTAAACAATCCAAACAACAAAAAACACATCTCTACACACACACACACAC 3652
QY 150 ATATACATACACACACACACACACACACACACATATTTTAACTCGGACATTAATGTA 209
DB 3651 ACACACACACACACACACACACACACACACACACACAAATTAAGTATTAATACATA 3592
QY 210 CACCACTATTCAAGTAAAGTAACTATCTCAGAGTAAAGACATATCTGCTATTGT 269
DB 3591 AACCTATATATCCCACTATTAAACAAAAACAAAAAATTAATAAAAAATA 3532
QY 270 CATTAAGTACGAGTTAA 286
DB 3531 CCTTAACAAAAAATTA 3515

```



















XX KY Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.  
 XX OS Homo sapiens.  
 XX PN WO2004060304-A2.  
 XX PD 22-JUL-2004.  
 XX PF 22-DEC-2003; 2003WO-US041389.  
 XX PR 27-DEC-2002; 2002US-0030773.  
 XX PA (SAGR-) SAGRES DISCOVERY INC.  
 XX PI Morris DW, Malandro MS;  
 XX DR WPI, 2004-543781/52.  
 XX PT New isolated cancer associated nucleic acids comprising at least 10  
 PT contiguous nucleotides, useful for diagnosing, preventing and/or treating  
 PT cancers such as leukemia and lymphoma.  
 XX PS Claim 1; SEQ ID NO 159; 199pp; English.  
 XX CC The present invention relates to cancer associated sequences (ADQ97025-  
 CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or  
 CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 138627 BP; 33083 A; 34745 C; 35023 G; 35281 T; 0 U; 495 Other;  
 Query Match 12.3%; Score 49; DB 12; Length 138627;  
 Best Local Similarity 82.1%; Pred. No. 0.013;  
 Matches 55; Conservative 1; Mismatches 11; Indels 0; Gaps 0;  
 OY 142 ATATATATATATATACACACACACACACACACACACATATTTAACTCGGACA 201  
 DB 53280 ATATATATATATACACACACACACACACACACACATATATATTTAGAGA 53221  
 OY 202 TAAATGT 208  
 DB 53220 TAAATTT 53214  
 RESULT 46  
 ABK28343/c  
 ID ABK28343 standard; DNA; 10138 BP.  
 XX ABK28343;  
 DT 23-APR-2002 (first entry)  
 XX DNA transcription associated genomic DNA #109.  
 XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KY PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KY single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KY viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KY immunological disorder; Werner syndrome; developmental disorder;  
 KY psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 KY neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;  
 KY myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KY angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
 KY polyglutamine disorder; solid tumour.  
 XX Unidentified.  
 XX WO200192565-A2.  
 XX 06-DEC-2001.

XX PF 06-APR-2001; 2001WO-EP003973.  
 XX PR 06-APR-2000; 2000DE-01019058.  
 XX PR 07-APR-2000; 2000DE-01019173.  
 XX PR 30-JUN-2000; 2000DE-01032529.  
 XX PR 01-SEP-2000; 2000DE-01043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI, 2002-090046/12.  
 XX PT New nucleic acids or oligomers, useful for diagnosing or treating  
 PT diseases associated with DNA transcription, e.g. immunological disorders,  
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or  
 PT cancer.  
 XX PS Claim 1; SEQ ID NO 217; 32pp; English.  
 XX CC The invention relates to a nucleic acid, which comprises a segment of the  
 CC chemically pretreated DNA of genes associated with DNA transcription from  
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 CC to the chemically pretreated DNA of genes associated with DNA  
 CC transcription. The set of oligomer probes are useful for detecting the  
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 CC diagnosing or treating diseases associated with DNA transcription  
 CC (particularly with the methylation status), e.g. adenosine deaminase  
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurological disorders, neurodegenerative disorders, Wardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification but  
 CC was obtained in electronic format directly from the European Patent  
 CC Office  
 XX SQ Sequence 10138 BP; 2605 A; 137 C; 2575 G; 4821 T; 0 U; 0 Other;  
 Query Match 12.2%; Score 48.8; DB 6; Length 10138;  
 Best Local Similarity 52.5%; Pred. No. 0.007;  
 Matches 127; Conservative 1; Mismatches 113; Indels 1; Gaps 1;  
 OY 142 ATATATATATATATACACACACACACACACACACACATATTTAACTCGGACA 201  
 DB 1181 AATATATACACACACACACACACACACACACACATATATATACCTTTAA 1122  
 OY 202 TAAATGTACACACCTTATCAAGTAAATATCTATCTGGAAGTAAAGCATATATCT 261  
 DB 1121 TCATCACCCCTTATCAATATATATATATATATATATATATATATATATATATAT 1062  
 OY 262 GCTATGTGATTAATCTAGC-AGTTAAGCTGCTTATGAGTTCACTTCCCTAC 320  
 DB 1061 TACCTTAAAAAATTTTATATATATATATATATATATATATATATATATATATAT 1002  
 OY 321 TATGAAAACTGAAAGGTGAAATCTGGTGTATTAATCTTAACTTTGAAAGCTCAATACC 380  
 DB 1001 AATTAATTTCAAAAATTTCAATCAATCAATTAATTAATAAATTTAAATTTACCTCA 942  
 OY 381 TA 382  
 DB 941 TA 940  
 RESULT 47  
 ACN43848



ID	ACN43848 standard; DNA; 40633 BP.
XX	AC
XX	ACN43848;
XX	18-NOV-2004 (first entry)
DT	
DE	Mouse genomic sequence MCG10049.
XX	
XX	Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
OS	Mus musculus.
XX	
PN	MO2003073826-A2.
XX	
PD	12-SEP-2003.
XX	
PF	28-FEB-2003; 2003WO-US006235.
XX	
PR	01-MAR-2002; 2002US-00087192.
XX	
PA	(SAGR-) SAGRES DISCOVERY.
PI	Morris DW;
DR	WPI; 2003-328604/31.
PT	
PS	Recombinant nucleic acid useful for diagnosis and treatment of carcinoma comprises a nucleotide sequence.
XX	
XX	Claim 1; SEQ ID NO 1; Opp; English.
CC	The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biologic; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US2002182586A1, for which no sequence data was published
SC	Sequence 40633 BP; 10555 A; 8795 C; 8589 G; 11134 T; 0 U; 1560 Other;
Query Match	12.2%; Score 48.8; DB 11; Length 40633;
Best Local Similarity	29.9%; Pred. No. 0.01;
Matches 59; Conservative 1; Mismatches 137; Indels 0; Gaps 0;	
OY	17 TATAATTTAGAGTCTTGGAATAAGCTTAATAATCCTGTGGCTACGATGAGACTGST 76 
DB	31067 TATAATATATANN 31126 
OY	77 GATACTGGTACTCTGGAAAGTAGATGCAGGCAAGTAGAGCCTACTTTTGTACT 136 NN 31186
OY	137 GTTGATATATATATATATACATACACACACACACACACACACATATTTACTCG 196 
DB	31187 NNN 31246 
OY	197 GGACATAAATGTACACC 213 
DB	31247 TTACTTTAATGATTAAC 31263 
RESULT 48	
ABD32966_01/c	
Continuation (2 of 11) of ABD32966 from base 100001 (Mouse cancer-associated genomic DN	
mp Sequence split into 11 fragments LOCUS ABD32966 Accession Abd32966	

WP	Fragment Name	Begin	End
WP	ABD32966_00	1	110000
WP	ABD32966_01	100001	210000
WP	ABD32966_02	200001	310000
WP	ABD32966_03	300001	410000
WP	ABD32966_04	400001	510000
WP	ABD32966_05	500001	610000
WP	ABD32966_06	600001	710000
WP	ABD32966_07	700001	810000
WP	ABD32966_08	800001	910000
WP	ABD32966_09	900001	1010000
WP	ABD32966_10	1000001	1089736

Query Match 12.2%; Score 48.8; DB 13; Length 110000;  
Best Local Similarity 66.7%; Pred. No. 0.014;  
Matches 68; Conservative 1; Mismatches 33; Indels 0; Gaps 0

QY 108 AGGGCAATGAGGCGCTACCTTTTGTACTGTTTGATATATATATATACATACACACAC 167  
DB 72370 ATGAAGGTGAGATTACCTCTCTGCAATATATGATACACACACACACACACACAC 723

QY 168 ACACACACACACACATATATTTTAACTCGGACATATATGTA 209  
DB 72310 ACACACACACACACACATATTTGTCTTATTTAAAAAGCA 72269

RESULT 49  
ABD33098/C  
ID ABD33098 standard; DNA; 219352 BP.  
XX ABD33098;  
AC  
XX  
XX  
DT 18-NOV-2004 (first entry)  
XX  
XX  
DE Murine cancer-associated (CA) gene MD07-008.  
XX  
XX  
KM Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
XX  
XX  
KW ds; cancer; cytostatic.  
XX  
XX  
OS Mus musculus.  
XX  
XX  
PN MO2004058146-A2.  
XX  
XX  
PD 15-JUL-2004.  
XX  
XX  
PF 15-DEC-2003; 2003MO-US040081.  
XX  
XX  
PR 17-DEC-2002; 2002US-00322281.  
XX  
XX  
PA (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX  
PI Morris DW, Malandro MS;  
XX  
XX  
DR WPI; 2004-499109/47.  
XX  
XX  
PT Novel human cancer associated protein encoded within open reading frame  
XX of cancer associated gene, useful as targets for diagnosing cancer.  
XX  
XX  
PS Disclosure; SEQ ID NO 45; 182bp; English.  
XX  
XX  
CC The invention relates to cancer-associated proteins (CAP) and the cancer-  
CC associated (CA) nucleic acids encoding them. The invention also relates  
CC to a method for treating cancers involving administering to a patient an  
CC inhibitor of CAP, and a method of screening for anticancer activity in a  
CC potential drug involving providing a cell that expresses a CA gene,  
CC contacting a tissue sample derived from a cancer cell with an anticancer  
CC drug candidate and monitoring the effect of the anticancer drug candidate  
CC on expression of the CA gene. The CAP proteins are useful for detecting  
CC cancer associated with expression of a CAP protein in a test cell sample  
CC and for screening for a bioactive agent capable of modulating the  
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing  
CC cancer, involving determining the expression of a CA nucleic acid in a  
CC tissue. This sequence represents a murine CA gene of the invention. Note:







**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2006, 16:57:45 ; Search time 35973 Seconds

(without alignments)  
621.792 Million cell updates/sec

Title: SEQ1-33670G  
Perfect score: 399.2  
Sequence: 1 aagaatgccttcctctcata.....taactcagatcagaagcgag 400

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues  
Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

EST:\*  
1: gb\_seq1:\*  
2: gb\_seq3:\*  
3: gb\_seq4:\*  
4: gb\_seq5:\*  
5: gb\_seq6:\*  
6: gb\_seq7:\*  
7: gb\_seq8:\*  
8: gb\_seq9:\*  
9: gb\_seq10:\*  
10: gb\_seq11:\*  
11: gb\_seq12:\*  
12: gb\_seq13:\*  
13: gb\_seq14:\*  
14: gb\_seq15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	217.6	54.5	681	4 BX486675	BX486675 DKF2P6860
2	57.6	14.4	882	14 CT350937	CT350937 Sus scrofa
3	54.2	13.6	1204	114 AG390615	AG390615 Mus muscu
4	53.8	13.5	655	4 CB940690	CB940690 IpcGJx14
5	53.6	13.4	868	14 CNO3132U	AL250131 Tetradon
6	53.6	13.4	612	14 FRO018554	AL011449 F. rubrip
7	53.6	13.4	779	9 DR106610	DR106610 Jhu118H03
8	53.4	13.4	491	11 AQ407117	AQ407117 HS_5076_A
9	53.4	13.4	600	11 B2217953	B2217953 CH230-518
10	53.4	13.3	757	10 DM252261	DM252261 UI-S-G80-
11	52.6	13.2	602	14 CT063242	CT063242 Sus scrofa
12	52.6	13.2	768	14 CT376921	CT376921 Sus scrofa
13	52.6	13.2	796	11 BH116000	BH116000 RPI-24-2
14	52.6	13.2	855	14 CT035778	CT035778 Sus scrofa
15	52.6	13.2	1015	12 CL121100	CL121100 ISB1-79P1
16	52.4	13.1	315	8 CN674247	CN674247 A0946A09-
17	52.4	13.1	521	8 CN676171	CN676171 A0976B03-
18	52.4	13.1	541	11 AQ353718	AQ353718 CITB1-E1-
19	52.4	13.1	619	11 AQ779979	AQ779979 HS_3113_A

20	52.4	13.1	736	14 AG561036	AG561036 Mus muscu
21	52.4	13.1	2558	6 AK142564	AK142564 Mus muscu
22	52.2	13.1	494	11 AQ209289	AQ209289 HS_3237_B
23	52.2	13.1	660	3 BQ188992	BQ188992 UI-E-E1-
24	52.2	13.1	665	3 BQ184399	BQ184399 UI-E-E1-
25	52.2	13.1	820	10 DM264376	DM264376 UI-S-GN1-
26	52.2	13.1	950	12 DU046739	DU046739 159296 To
27	52	13.0	448	11 A0868109	A0868109 nbdb0021B
28	52	13.0	799	14 CR023958	CR023958 Forward B
29	52	13.0	887	13 CD360167	CD360167 AGENCOURT
30	51.8	13.0	345	13 CL440749	CL440749 PSTVU01.H
31	51.8	13.0	501	2 B1503732	B1503732 B170022A
32	51.8	13.0	649	13 CM904996	CM904996 RPI142.14
33	51.6	12.9	528	4 CB940280	CB940280 IpcGJx14
34	51.6	12.9	694	14 BX162787	BX162787 Dario rer
35	51.6	12.9	827	14 BX128732	BX128732 Dario rer
36	51.4	12.9	528	11 BH343871	BH343871 CH230-41N
37	51.4	12.9	777	11 B2250267	B2250267 CH230-353
38	51.2	12.8	178	11 A2514487	A2514487 IMC361P24
39	51.2	12.8	541	11 AZ703674	AZ703674 RPI-23-2
40	51.2	12.8	750	14 CT454907	CT454907 Sus scrofa
41	51	12.8	607	10 DM340951	DM340951 PE_LEC000
42	51	12.8	607	11 AQ326274	AQ326274 nbdb0028M
43	51	12.8	652	9 DN877029	DN877029 nae10901.
44	51	12.8	672	11 AQ923572	AQ923572 RPI-23-2
45	51	12.8	835	14 CT398920	CT398920 Sus scrofa
46	51	12.8	1072	10 DM594541	DM594541 CGX115-F0
47	51	12.8	1381	9 DN738453	DN738453 CNB98-F11
48	51	12.8	1722	6 AK139565	AK139565 Mus muscu
49	50.8	12.7	469	14 DE117100	DE117100 Oryzias 1
50	50.8	12.7	493	11 AZ073117	AZ073117 RPI-23-4
51	50.8	12.7	599	11 AQ262841	AQ262841 CITB1-E1-
52	50.8	12.7	714	14 BX176458	BX176458 Dario rer
53	50.8	12.7	732	13 DU341743	DU341743 109831304
54	50.8	12.7	770	14 CR830901	CR830901 GROMAA63B
55	50.8	12.7	879	13 CZ992356	CZ992356 209924 To
56	50.8	12.7	1175	4 BX363884	BX363884 BX363884
57	50.6	12.7	488	11 AQ597715	AQ597715 HS_2090_B
58	50.6	12.7	590	11 AZ012541	AZ012541 RPI-23-3
59	50.6	12.7	693	14 AG171472	AG171472 Pan trogl
60	50.6	12.7	825	14 CT053907	CT053907 Sus scrofa
61	50.6	12.7	892	14 CT354182	CT354182 Sus scrofa
62	50.4	12.6	274	3 BP510846	BP510846 BP510846
63	50.4	12.6	525	11 B64703	B64703 CIT-HSP-202
64	50.4	12.6	606	14 DR36111T	DR36111T Dario rer
65	50.4	12.6	607	14 CR880635	CR880635 Sus scrofa
66	50.4	12.6	623	14 BX980437	BX980437 Forward S
67	50.4	12.6	711	11 AZ337912	AZ337912 IM0068G18
68	50.4	12.6	780	14 CT048379	CT048379 Sus scrofa
69	50.4	12.6	2533	6 AK032764	AK032764 Mus muscu
70	50.4	12.6	5305	6 AK143960	AK143960 Mus muscu
71	50.4	12.6	540	13 CM353832	CM353832 fdbb001f0
72	50.2	12.6	608	10 DT448882	DT448882 JGI_CABK8
73	50.2	12.6	633	10 DT448883	DT448883 JGI_CABK8
74	50.2	12.6	674	14 AG034856	AG034856 Pan trogl
75	50.2	12.6	835	3 CX823483	CX823483 JGI_CAK3
76	50.2	12.6	1001	3 BQ941666	BQ941666 AGENCOURT
77	50	12.5	274	11 AZ517401	AZ517401 RPI-11-2
78	50	12.5	617	12 CG976896	CG976896 CH240.167
79	50	12.5	619	11 AQ342672	AQ342672 RPI11-11
80	50	12.5	732	13 CM566830	CM566830 OA_ABA011
81	50	12.5	766	14 AG496779	AG496779 Mus muscu
82	50	12.5	814	14 CC905634	CC905634 t024r22ba
83	50	12.5	815	12 CT319339	CT319339 Sus scrofa
84	50	12.5	933	10 DM662837	DM662837 CNB316-GO
85	50	12.5	952	13 CZ941129	CZ941129 323146 To
86	49.8	12.5	563	12 BZ599380	BZ599380 WRADP43TF
87	49.8	12.5	634	14 BX157385	BX157385 Dario rer
88	49.8	12.5	646	14 AG065044	AG065044 Pan trogl
89	49.8	12.5	750	3 BQ002660	BQ002660 UI-H-E11-
90	49.8	12.5	751	4 AG603288	AG603288 Mus muscu
91	49.8	12.5	763	13 CL987558	CL987558 ZMWBE000
92	49.8	12.5	787	10 DT292845	DT292845 JGI_CAW4



C 93	49.8	12.5	803	13	CZ979367	CZ979367 197206 To
C 94	49.6	12.4	468	1	AL709462	AL709462 DKFZp686D
C 95	49.6	12.4	500	7	BH296556	BH296556 601174057
C 96	49.6	12.4	527	1	AJ424934	AJ424934 AJ424934
C 97	49.6	12.4	550	14	CR223106	CR223106 Forward s
C 98	49.6	12.4	586	14	CR260017	CR260017 Forward s
C 99	49.6	12.4	588	11	AQ211188	AQ211188 HS 5564 A
C 100	49.6	12.4	626	11	BZ214131	BZ214131 CR230-426
C 101	49.6	12.4	638	14	CR117368	CR117368 Forward s
C 102	49.6	12.4	662	14	DU655930	DU655930 OG AB8005
C 103	49.6	12.4	702	14	CR158437	CR158437 Forward s
C 104	49.6	12.4	758	14	AG526268	AG526268 Mus muscu
C 105	49.6	12.4	811	14	DU860221	DU860221 66448 Tom
C 106	49.6	12.4	829	14	CT406476	CT406476 Sus scrofa
C 107	49.6	12.4	905	14	CR235414	CR235414 Forward s
C 108	49.6	12.4	239	14	DE236382	DE236382 Tifolium
C 109	49.4	12.4	258	14	BX222469	BX222469 Danio rer
C 110	49.4	12.4	267	11	AZ121960	AZ121960 RPCI-23-1
C 111	49.4	12.4	290	3	BQ989052	BQ989052 OGF16024
C 112	49.4	12.4	292	14	CT427657	CT427657 Sus scrofa
C 113	49.4	12.4	377	12	BZ930241	BZ930241 CH240.35T
C 114	49.4	12.4	481	7	BF703922	BF703922 M1-P-A5-a
C 115	49.4	12.4	532	9	DN871044	DN871044 nad1.0N06
C 116	49.4	12.4	561	10	DV722099	DV722099 RYL13435
C 117	49.4	12.4	570	11	BH084190	BH084190 RPCI-24-3
C 118	49.4	12.4	572	12	CE736719	CE736719 L19T-988-
C 119	49.4	12.4	626	4	BX687247	BX687247 BX687247
C 120	49.4	12.4	632	14	BX128212	BX128212 Danio rer
C 121	49.4	12.4	656	12	CE784742	CE784742 L19T-988-
C 122	49.4	12.4	662	9	DN744924	DN744924 GL-CI-171
C 123	49.4	12.4	763	14	CT210091	CT210091 Sus scrofa
C 124	49.4	12.4	781	14	CT189540	CT189540 Sus scrofa
C 125	49.4	12.4	789	11	CT101972	CT101972 Sus scrofa
C 126	49.4	12.4	860	11	BZ274753	BZ274753 CH230-450
C 127	49.4	12.4	947	14	CNS02MB6	CNS02MB6 Tetracton
C 128	49.4	12.4	981	12	CL054028	CL054028 CH216-78K
C 129	49.2	12.3	223	12	CE396292	CE396292 L19T-988-
C 130	49.2	12.3	269	7	BE501814	BE501814 hms5ell.x
C 131	49.2	12.3	377	10	DM451534	DM451534 HRA50042
C 132	49.2	12.3	435	11	AQ041986	AQ041986 CTT-HSP-2
C 133	49.2	12.3	531	12	CL193434	CL193434 104_417-1
C 134	49.2	12.3	540	11	AZ952702	AZ952702 2M021E11
C 135	49.2	12.3	658	11	AZ643106	AZ643106 IM0506L05
C 136	49.2	12.3	702	14	CR048186	CR048186 Forward s
C 137	49.2	12.3	746	14	AG606779	AG606779 Mus muscu
C 138	49.2	12.3	753	11	AQ895904	AQ895904 HS_3176_B
C 139	49.2	12.3	777	11	AQ742337	AQ742337 HS_5383_B
C 140	49.2	12.3	780	14	CR267037	CR267037 Reverse s
C 141	49.2	12.3	783	14	BX971441	BX971441 Reverse s
C 142	49.2	12.3	785	14	AG349596	AG349596 Mus muscu
C 143	49.2	12.3	808	14	CR007296	CR007296 Reverse s
C 144	49	12.3	349	4	BX955892	BX955892 DKFZp7811
C 145	49	12.3	564	11	AZ329878	AZ329878 IM0054L08
C 146	49	12.3	564	11	AZ515186	AZ515186 IM0054L08
C 147	49	12.3	589	13	DUI175448	DUI175448 109855324
C 148	49	12.3	653	11	AZ603140	AZ603140 IM0422G17
C 149	49	12.3	725	14	AG497075	AG497075 Mus muscu
C 150	49	12.3	762	14	AG399545	AG399545 Mus muscu

## ALIGNMENTS

RESULT 1  
LOCUS BX486675 681 bp mRNA linear EST 04-SEP-2003  
DEFINITION DKFZp686006252.1 686 (synonym: hlc3) Homo sapiens cDNA clone  
ACCESSION BX486675  
VERSION BX486675.1 GI:31950573  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Bahr, A., Lauber, J., Mewes, H.W., Weill, B., Amid, C., Oanger, A., Fodor, G., Han, W., and Wiemann, S.  
TITLE EST (Bah, A., Lauber, J., Mewes, H.W., Weill, B., et al.)  
JOURNAL Unpublished (2003)  
COMMENT Contact: MIPS

FEATURES  
source  
1..681  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686006252"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_id="686 (synonym: hlc3)"  
/note="Vector: pTriblex2; Site\_1: SfilA; Site\_2: SfilB; DNA-collection"

## ORIGIN

Query Match 54.5%; Score 217.6; DB 4; Length 681;  
Best Local Similarity 96.8%; Pred. No. 2.6e-44;  
Matches 242; Conservative 2; Mismatches 1; Indels 5; Gaps 2;

1 AAGATGCTCTTCTCTATATATTACGAGTCTTGATAGCTTAATAATACCTGTGGC 60  
249 AAGATGCTCTTCTCTATATATTACGAGTCTTGATAGCTTAATAATACCTGTGGC 190  
Db  
61 TACACATGACGTGTGATACCTGTTACCTCTGGAAGTATGACGAGGCAAGTACG 120  
189 TACACATGAACTGTGATACCTGTTACCTCTGGAAGTATGACGAGGCAAGTACG 130  
Db  
121 CTTACCTTTTGTACTGTTGATATATATATATATATATATATATATATATATAT 176  
129 CTTACCTTTTGTACTGTTGATATATATATATATATATATATATATATATATAT 70  
Db  
177 CACACATATATTTTACCTCGGACATATATATATATATATATATATATATATATAT 236  
69 CACACATATATTTTACCTCGGACATATATATATATATATATATATATATATATAT 11  
Db  
237 ATTCGAGAG 246  
Db  
10 ATTCGAGAG 1

RESULT 2  
LOCUS CT350937 882 bp DNA linear GSS 03-NOV-2005  
DEFINITION SUE SCROFA genomic clone CH242-413G22, genomic survey sequence.  
ACCESSION CT350937  
VERSION CT350937.1 GI:79926432  
KEYWORDS GSS.  
SOURCE SUE SCROFA (pig)  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae; Suidae.  
REFERENCE 1 (bases 1 to 882)  
AUTHORS Humphrey, S.J., Plumb, R.W. and Durham, J.L.  
TITLE Direct Submission



**JOURNAL** Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [hunquerry@sanger.ac.uk](mailto:hunquerry@sanger.ac.uk) Unpublished

**COMMENT** This sequence was generated from the SPE end of BAC 413G22. 413G22 is part of the CHORI-242 BAC Library created by P. de Jong. Further details: [http://www.sanger.ac.uk/Projects/S\\_sacrofa/](http://www.sanger.ac.uk/Projects/S_sacrofa/).

**FEATURES** Location/Qualifiers

source

1..882

/organism="Sus scrofa"

/mol\_type="genomic DNA"

/db\_xref="taxon:9823"

/clone="CH242-413G22"

/tissue\_type="white blood cells"

/note="vector pTARBAC1.3\_BamHI sex female"

**ORIGIN**

Query Match 14.4%; Score 57.6; DB 14; Length 882;  
 Best Local Similarity 55.9%; Pred. No. 0.00084;  
 Matches 105; Conservative 2; Mismatches 81; Indels 0; Gaps 0;

4 AATGCTTTCTCATATATTACGAGTTCTTGAAATAGCTTAATACTGTGGCGTAC 63  
 |||||  
 315 AATCTCTGCTTATATCTTAACTTTTATTAATAAAGATATCTATTTAGT 374  
 |||||  
 64 ACATGAGACTGCTGCTGCTTACCTCTGGAAGTAGATGACGAGGAGTGGGCT 123  
 |||||  
 375 TGAATTAAGAGTTAGTTGCTGCTATATAGCAACACATTCAGTTACATTAAT 434  
 |||||  
 124 ACCCTTTTGTACTGTTGATATATATATACATACACACACACACACACACAY 183  
 |||||  
 435 ACATATATATTCATTTATATATATATATATATACACACACACACACACACAC 494  
 |||||  
 184 ATATTTTA 191  
 |||||  
 495 ACATATTA 502  
 |||||

**RESULT 3** AG390615/c 1204 bp DNA linear GSS 21-DEC-2004  
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-207F19.T7, genomic survey  
 DEFINITION AG390615  
 ACCESSION AG390615  
 VERSION AG390615.1 GI:48008031  
 KEYWORDS GSS.  
 SOURCE Mus musculus molossinus (Japanese wild mouse)  
 ORGANISM Mus musculus molossinus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Abe, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T., Kawata, K., Saitou, N., Hattori, M., Sakaki, Y., Moriaki, K. and Shiroishi, T.  
 Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis  
 JOURNAL Genome Res. 14 (12), 2439-2447 (2004)  
 PUBMED 15574823  
 REFERENCE 2 (bases 1 to 1204)  
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
 JOURNAL Direct Submission  
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shohji-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: [hattori@gscc.riken.jp](mailto:hattori@gscc.riken.jp), URL: <http://ngp-gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe ([abe@gscc.riken.jp](mailto:abe@gscc.riken.jp)). The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199  
 e-mail: [abe@gscc.riken.jp](mailto:abe@gscc.riken.jp)

**PRIMERS** Sequencing : T7  
 LIBRARY : PBACe3.6  
 Vector : EcoRI  
 R Site 1 : EcoRI  
 R Site 2 : EcoRI.

**FEATURES** Location/Qualifiers

source

1..1204

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-207F19.T7"

/sex="male"

/tissue\_type="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

**ORIGIN**

Query Match 13.6%; Score 54.2; DB 14; Length 1204;  
 Best Local Similarity 76.5%; Pred. No. 0.0062;  
 Matches 65; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

123 TAACCTTTTGTACTGTTGATATATATATATATACATACACACACACACACACA 182  
 |||||  
 374 TGCTTTGCTTACGTTGATATATATATATATACATACATACATACACACACACATCA 315  
 |||||  
 183 YATATTTTAACCTGGGACATTAATG 207  
 |||||  
 314 TACATATTGCTCTGAATTCGAAG 290  
 |||||

**RESULT 4** CB940690/c 655 bp mRNA linear EST 29-APR-2003  
 LOCUS IPEGJx14.22.H05.23 IPEGJx14 Ictalurus punctatus cDNA clone  
 DEFINITION IPEGJx14.22.H05.23, mRNA sequence.  
 ACCESSION CB940690  
 VERSION CB940690.1 GI:30226081  
 KEYWORDS EST.  
 SOURCE Ictalurus punctatus (channel catfish)  
 ORGANISM Ictalurus punctatus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Siluriformes; Ictaluridae; Ictalurus.  
 1 (bases 1 to 655)  
 Bengten, R., Khayat, M., Middleton, D., Waldbieser, G., Askovic, S., Jensen, K.T., Marr, G., Miller, N., Clem, L.W. and Wilson, M.  
 Identification of expressed genes in mixed leukocyte culture and in a macrophage cell line in channel catfish, Ictalurus punctatus  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Waldbieser GC  
 Catfish Genetics Research Unit  
 USDA-Agricultural Research Service  
 141 Experiment Station Road, Stoneville, MS 38776, USA  
 Tel: 662 686 3593  
 Fax: 662 686 3567  
 Email: [gwaldbieser@ars.ubda.gov](mailto:gwaldbieser@ars.ubda.gov)  
 Single pass sequencing. Bases called with Phred v0.000925.c. Low quality bases and vector trimmed with Lucy v1.16.  
 Plate: 22 row: H column: 5  
 Seq primer: T7.  
 Location/Qualifiers

1..655

/organism="Ictalurus punctatus"

/mol\_type="mRNA"

/strain="commercial"

/db\_xref="taxon:7998"

/clone="IPEGJx14.22.H05"

/cell\_type="macrophage"

/cell\_line="427A"

/dev\_stage="adult"



```

ORIGIN
      /lab_host="DH108"
      /clone_id="IpCGx14"
      /note="Vector: pSPORT1, Site_1: Sal 1, Site_2: Not 1,
      Primary library"

```

Query Match	13.5%	Score 53.8	DB 4	length 655
Best Local Similarity	52.0%	Pred. No. 0.0075		
Matches 118; Conservative	1	Mismatches 108	Indels 0	Gaps 0

[illegible]

RESULT 5			
CNS03LZU/c	LOCUS	DEFINITION	
CNS03LZU	868 bp	DNA	linear
Tetradon nigroviridis genome survey sequence PUC-ORI end of clone 03604 of library G from Tetradon nigroviridis, genomic survey sequence.			

ACCESSION AL250131  
VERSION AL250131.1 GI:7971143  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetradodon nigroviridis  
ORGANISM Tetradodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorphi; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodontiformes;

1  
**REFERENCE**  
**AUTHORS** Roset Croillien,H., Jallion,O., Desliva,C., Bouneau,L., Fisher,C.,...  
 Benoit,A., Fzmes,C., Wincker,P., Brotier,P., Quetier,F.,  
 Saurin,W. and Weissenbach,J.  
**TITLE** Estimate of human gene number provided by genome-wide analysis  
 using Tetradon nigroviridis DNA sequence  
**JOURNAL** Nat. Genet. 25 (2), 235-238 (2000)  
**PUBMED** 10635645

AUTHORS	TITLE	JOURNAL	PUBLISHED	REFERENCE
Rosset Crollius, H., Jailion, O., Dasilva, C., Orou-Cottar, C., Flamenc, C., Flecher, C., Bonnaud, L., Billault, A., Quettar, F., Saurin, W., Bernot, A. and Weissenbach, J.	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Genome Res.	10 (7), 939-949 (2000)	10899143
				3 (bases 1 to 868)

**JOURNAL** Submitted (12-APR-2000) Genoscope - Centre National de Séquençage : BP 191 91006 Evry cedex - FRANCE (E-mail : [segretefgenoscope.cns.fr](mailto:segretefgenoscope.cns.fr) - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

**COMMENT** This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.

```

FEATURES
    source
    location/Qualifiers
    1. .868
    /organism="Tetradodon nigroviridis"
    /mol_type="genomic DNA"
    /db_xref="taxon:99983"

```

```

ORIGIN
      /clone="036004"
      /clone_lib="G"
      /note="Genoscope sequence ID : COBG036DE02SP1
end : PUC-Orig"

```

Query Match	13.5%	Score 53.8	DB 14	Length 868
Best Local Similarity	76.5%	Pred. NC. 0.0077		
Matches 62, Conservative	3	Mismatches 16	Indels 0	Gaps 0

QY  
131 TGTACTGTTTGAATATATATATATACATCACACACACACACACACACATATTTTT 190

DB  
135 TATACATATATATATATATATATATACACACACACACACACACACATATATATA 76

QY  
191 AACTCGGGACATTAATGTACA 211

DB  
75 TAATCTGTAATATATATATATATA 55

RESULT 6		
PRO018554		
LOCUS	F rubripes	612 bp DNA, linear
DEFINITION	GSS sequence, clone O16P03aC3,	GSS 25-PB8-2004
ACCESSION	AL011449	genomic survey sequence.
VERSION	AL011449.1	
KEYWORDS	GI:2676882	
SOURCE	GSS; genome survey sequence.	
ORGANISM	Takifugu rubripes (Fugu rubripes)	
	Takifugu rubripes	

REFERENCE

1. Eulaxyotidae, Metastoda; Chordata, Crustacea; Vertebrates; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Buteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.

AUTHORS	TITLE
Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Edwards, Y. J., Bouchireb, N., Cottage, A., Yeo, G.S., Umranta, Y., Williams, G. and Brenner, S.	Generation and analysis of 25 Mb of genomic DNA from the pufferfish <i>Fugu rubripes</i> by sequence scanning
JOURNAL PUBLISHED	Genome Res. 9 (10), 960-971 (1999)
REFERENCE	10523524
AUTHORS	2 (bases 1 to 612)
	Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,

TITLE  
Journal  
COMMENT  
V. type: phage  
PRIMER: KS  
DESCR:

Williams, G. and Brenner, S.  
Direct Submission  
Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB3 0SB. Email: [biohel@hgm.mrc.ac.uk](mailto:biohel@hgm.mrc.ac.uk)  
Vector: pBlueScript II KS  
V. type: phage  
PRIMER: KS  
DESCR:

```

sequence.
location/Qualifiers
1..612
  /organism="Takifugu rubripes"
  /mol_type="genomic DNA"
  /db_xref="taxon:31033"
  /clone="016P03ac3"
  /clone_id="cosmid_016P03"

```

Query Match	13.4%	Score 53.6	DB 14	Length 612
Best Local Similarity	60.2%	Pred. No. 0.0084		
Matches	77	Conservative	1	Mismatches 50; Indels 0; Gaps 0

  

QY	133	TACTGTTGATATATATATATACATACACACACACACACATATATTTTAA	192
DB	132	TCCTGTTTCATATAAAGAAACACACACACACACACACACACACACACACA	191

OY  
Db

193 CTCGGACATAATGTACACCACCTATTGAAAGTAAAACTACTATTCGAAGTAAGA 252  
| | | | | | | | | | | | | | |  
192 CMCANTTGTTAAGNAAATATNNCGAGTTNAAGTAAATACMTCCTACTGCANGAGAAA 251



QY 253 CATATCC 260  
 |||||  
 DB 252 GATATCC 259

RESULT 7  
 LOCUS DR106610  
 DEFINITION JH118H03L779 Canine cardiovascular system biased cDNA Canis familiaris cDNA similar to He cell division cycle associated 2 (CCCA2), mRNA sequence.  
 ACCESSION DR106610  
 VERSION DR106610  
 KEYWORDS GI:67565965  
 SOURCE EST.  
 ORGANISM Canis familiaris (dog)  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 779)  
 AUTHORS Disilverio, D., Yung, C., Gao, Z., Farukhi, Y., Winslow, R.L. and Tomaseili, G.F.  
 TITLE Canine cardiovascular system biased cDNA sequences  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: Gordon F. Tomaseili  
 Johns Hopkins University  
 720 Rutland Avenue/Koss 844, Baltimore, MD 21205, USA  
 Tel: 4109552774  
 Fax: 4105022096  
 Email: gtomaseil@jhmi.edu.

FEATURES  
 source location/Qualifiers  
 1..779  
 /organism="Canis familiaris"  
 /mol\_type="mRNA"  
 /strain="Mixed"  
 /db\_xref="taxon:9615"  
 /sex="Mixed"  
 /clone\_lib="Canine cardiovascular system biased cDNA"  
 /note="Organ: Mixed; Vector: pCDNA3.1; Site\_1: EcoRI; Site\_2: XhoI; Adult tissue from eye, lung, aorta, pulmonary artery and brain; neonatal tissue from liver, spleen, thymus, lung, kidney, aorta and brain; 50% is from cardiac tissues."

ORIGIN  
 Query Match 13.4%; Score 53.6; DB 9; Length 779;  
 Best Local Similarity 67.3%; Pred. No. 0.0086;  
 Matches 74; Conservative 1; Mismatches 35; Indels 0; Gaps 0;

QY 122 CTACCTTTTGTACTGTTGATATATATATATATACACACACACACACACAC 181  
 |||||  
 DB 422 CTCTCTCTCTCTCTCTCTATATATATATATATACACACACACACACAC 481

QY 182 AATATATTTTAACTCGGGACATAATATGACACACCTTTCAAGTAAATAA 231  
 |||||  
 DB 482 ACATATATACATATACCATGATTAATATATATATATATATATATATAT 531

RESULT 8  
 LOCUS AQ407117/c 491 bp DNA linear GSS 17-MAR-1999  
 DEFINITION HS\_5076\_A2\_H04\_SP68 RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=652 Col=8 Row=O, genomic survey sequence.  
 ACCESSION AQ407117  
 VERSION AQ407117.1 GI:4429739  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 491)  
 AUTHORS Mahatras, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,

TITLE Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.  
 JOURNAL Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 PUBMED Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 COMMENT 10449764  
 Contact: Mahatras GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 652 row: O column: 8  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 491.

FEATURES  
 source location/Qualifiers  
 1..491  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=652 Col=8 Row=O"  
 /sex="male"  
 /clone\_lib="RPCT-11 Human Male BAC library"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

ORIGIN  
 Query Match 13.4%; Score 53.4; DB 11; Length 491;  
 Best Local Similarity 74.2%; Pred. No. 0.0093;  
 Matches 66; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 136 TGTTCAT 195  
 |||||  
 DB 227 TGTTCAT 168

QY 196 GGGACATAATATGACACACCTTATTCAA 224  
 |||||  
 DB 167 AT 139

RESULT 9  
 LOCUS BZ217953 600 bp DNA linear GSS 11-OCT-2002  
 DEFINITION CH230-51817.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone  
 ACCESSION BZ217953  
 VERSION BZ217953.1 GI:23876311  
 KEYWORDS GSS.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 600)  
 AUTHORS Zhao, S., Shetty, J., Shatman, S., Teagave, G., Geer, K., Shwartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.  
 TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other GSSs: CH230-51817.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics







REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
ORIGIN

1 (bases 1 to 602)  
Humphray,S.J., Plumb,R.W. and Durham,J.L.  
Direct Submission  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humphray@sanger.ac.uk Unpublished  
This sequence was generated from the T7 end of BAC 51E11. 51E11 is  
part of the PIGBAC BAC library created by Roslin Institute/RFCGR.  
Further details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/)  
location/Qualifiers  
1. 602  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="PigB-51E11"  
/tissue\_type="blood cells"  
/note="Vector pBeloBAC11  
sex male"

Query Match 13.2% Score 52.6; DB 14; Length 602;  
Best Local Similarity 67.0%; Pred. No. 0.015; Indels 0; Gaps 0;  
Matches 73; Conservative 1; Mismatches 35; Indels 0; Gaps 0;

ORIGIN

Query Match 13.2% Score 52.6; DB 14; Length 768;  
Best Local Similarity 63.2%; Pred. No. 0.015; Indels 0; Gaps 0;  
Matches 79; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

142 ATATATATATATATACATACACACACACACACACACACACACATATTTTAACTCGGAGCA 201

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
ORIGIN

1 (bases 1 to 768)  
Humphray,S.J., Plumb,R.W. and Durham,J.L.  
Direct Submission  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humphray@sanger.ac.uk Unpublished  
This sequence was generated from the SP6 end of BAC 445B16. 445B16  
is part of the CHORI-242 BAC library created by P. de Jong. Further  
details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/)  
location/Qualifiers  
1. 768  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-445B16"  
/tissue\_type="white blood cells"  
/note="Vector pTARBAC1.3\_BamHI  
sex female"

Query Match 13.2% Score 52.6; DB 14; Length 768;  
Best Local Similarity 63.2%; Pred. No. 0.015; Indels 0; Gaps 0;  
Matches 79; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

142 ATATATATATATATACATACACACACACACACACACACACATATTTTAACTCGGAGCA 201

Db	620	ATATTATTAATAATGACACAGCACACACACACACACACATCACTACGCCTTCCATG	
Qy	202	TAAATGTACACCACCTATTTCGAAGTAAAAACTATCTCAGAGAAGTAAAGACATATCTCT	261
Db	680	TATATATACACAGCATATATATGTGTATATATATACACCCACATATATATACTATACATA	739
Qy	262	GCTAT 266	
Db	740	TATAT 744	
RESULT 13			
BH116000                      796 bp    DNA       linear    GSS 19-JUL-2000			
LOCUS                          RPCI-24-253C11.TV RPCI-24 Mus musculus genomic clone			
DEFINITION                   RPCI-24-253C11, genomic survey sequence.			
ACCESSION                   BH116000			
VERSION                     BH116000.1 GI:14955795			
KEYWORDS                    GSS.			
SOURCE                      Mus musculus (house mouse)			
ORGANISM                    Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;			
Sciurognathi; Murioidea; Muridae; Murinae; Mus.			
REFERENCE			
AUTHORS                    Zhao,S., Niemman,W., Malek,J., Shatsman,S., Aktinret,B., Levins,M.,			
Tsengaye,G., Geer,K., Krol,M., Shvartabeyn,A., Gebregeorgis,E.,			
Russell,D., de Jong,P. and Fraser,C.M.			
Mouse BAC End Sequences from Library RPCI-24			
Unpublished (1999)			
TITLE			
JOURNAL                     Other_GSSs: RPCI-24-253C11.TV			
COMMENT                    Contact: Shaying Zhao			
Department of Eukaryotic Genomics			
The Institute for Genomic Research			
9712 Medical Center Dr., Rockville, MD 20850, USA			
Tel.: 301 838 0200			
Fax: 301 838 0208			
Email: szhao@igr.org			
Clones are derived from the mouse BAC library RPCI-24. For BAC			
library availability, please contact Pieter de Jong			
(pje@omni.net). Clones may be purchased from BACPAC			
Resources ( <a href="http://www.chori.org/bacpac/orderingframe.htm">http://www.chori.org/bacpac/orderingframe.htm</a> ). BAC end			
page: <a href="http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html">http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html</a>			
Plate: 253 row: C column: 11			
Seq primer: T7			
Class: BAC ends.			
FEATURES			
SOURCE			
Location/Qualifiers			
1..796			
/organism="Mus musculus"			
/mol_type="genomic DNA"			
/strain="C57BL/6J"			
/db_xref="taxon:10090"			
/clone="RPCI-24-253C11"			
/sex="Male"			
/cell_type="Spleen/Brain"			
/clone_id="RPCI-24"			
/note="Vector: pTRABAC1, Site_1: BamH1, Site_2: BamH1,			
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The			
library was cloned in the pTRABAC1 cloning vector at the			
BamH1 sites using MboI partially digested male C57BL/6J			
DNA."			
ORIGIN			
Query Match                   13.2%; Score 52.6; DB 11; Length 796;			
Best Local Similarity        65.0%; Pred. NO. 0.015;			
Matches                      76; Conservative 1; Mismatches 40; Indels 0; Gaps 0			
Qy	118	AGGCTACCTTTTGTACTGGTAGTATATATATATATACATACACACACACACACAC	177
Db	341	ATGCACATATTTATACACATCATCTTCATATATATATATATATACACACACACACACAC	400
Qy	178	ACAACATATTTTAACTCGGAGACATTAATGTACACCACTTATCAAGTAAAAGTA	234



Db: 401 ACACACACATATATACCATACACACAAACACACATATCTCAACACACACCA 457

RESULT 14  
CT035778 855 bp DNA linear GSS 01-NOV-2005  
LOCUS Sns scrofa genomic clone CH242-1P14, genomic survey sequence.  
DEFINITION CT035778  
ACCESSION CT035778.1 GI:78582246  
KEYWORDS GSS.  
SOURCE Sns scrofa (pig)  
ORGANISM Sns scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE  
1 (bases 1 to 855)  
Humphray, S.J., Plumb, R.W. and Durham, J.L.  
Direct Submission  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished  
This sequence was generated from the 77 end of BAC 1P14. 1P14 is part of the CHORI-242 BAC library created by P. de Jong. Further details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/).

FEATURES  
source  
1..855  
/organism="Sns scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-1P14"  
/issue\_type="White blood cells"  
/note="vector pTRABAC1.3\_BamHI  
sex female"

ORIGIN  
Query Match 13.2%; Score 52.6; DB 14; Length 855;  
Best Local Similarity 56.3%; Pred. No. 0.015;  
Matches 94; Conservative 2; Mismatches 71; Indels 0; Gaps 0;

QY 20 AATATTGCGAGTTTGAATTAAGTTAAATACCTGTGGCGTACATGAGCTGAT 79  
DB 241 AATATGAATGAGCTCAACAAATGTTCTCTCTGTGATGAGTTATTTTCATTTGCAT 182  
QY 80 ACAGGTATACCTCGGAAGTAGTAGGAGGCAAGGAGGCGTACCTTTTGTACAGTT 139  
DB 181 AATGCCCTCAGGTTCTATCCAGTTGTCAAAATGGCAAGATATATCTTTCTCATGGC 122  
QY 140 TGAATATATATATATACATACACACACACACACACACACACAYATA 186  
DB 121 TGAATATACACACACACACACACACACACACACACACACACACA 75

RESULT 15  
CL211100 1015 bp DNA linear GSS 05-JAN-2004  
LOCUS ISB1-79P10 Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-79P10,  
DEFINITION genomic survey sequence.  
ACCESSION CL211100  
VERSION CL211100.1 GI:40614735  
KEYWORDS GSS.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE  
1 (bases 1 to 1015)  
Kemnitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.  
A physical map of the xenopus tropicalis genome  
Unpublished (2003)  
Contact: Richard K Wilson  
Genome Sequencing Center

Washington University School of Medicine  
Email: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
Insert Length: 75000 Std Error: 0.00  
Seq primer: Sp6 ATTTAGTGCACACTTAG  
Class: BAC ends  
High quality sequence start: 12  
High quality sequence stop: 766.  
Location/Qualifiers  
1..1015  
/organism="Xenopus tropicalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:8364"  
/clone="ISB1-79P10"  
/note="ISB1"  
/note="vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC  
Library Segment 1"

ORIGIN  
Query Match 13.2%; Score 52.6; DB 12; Length 1015;  
Best Local Similarity 57.0%; Pred. No. 0.016;  
Matches 94; Conservative 1; Mismatches 70; Indels 0; Gaps 0;

QY 138 TTGTATATATATATATATATACACACACACACACACACACATATTTAACTCGG 197  
DB 214 TATATATATATATATATATATATATATATACACACACACACACACATATTTCAAGCTTCT 273  
QY 198 GACATTAAGTAGACACACACCATTCAAAGTAAACTACTCTCAGAGTAAGACATAA 257  
DB 274 TTCAATATTTTAAGCTTAATATATATATCAAACTTTCATATGTTGAGTAAGAGAAAA 333  
QY 258 TCGTCTATTTGTCATTAACCTAGCAAGTTAACTGCTTAACCTTAT 302  
DB 334 CAATATCCTGCAATGTAAATGTAGAAACAAATATTAATTTCT 378

RESULT 16  
CN674247 315 bp mRNA linear EST 17-MAY-2004  
LOCUS A0946A09-5 NIA Mouse Embryonic Stem (ES) cell (Lif+ 48 h, high  
DEFINITION density) cDNA library (Long) Mus musculus cDNA clone NIA:A0946A09  
IMAGE:30771176 5', mRNA sequence.  
CN674247  
CN674247.1 GI:47440698  
EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 315)  
Sharov, A.A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y., Wang, Y., Carter, M.G., Hamatani, T., Albe, K., Akutsu, H., Sharov, L., Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S., Nagaraja, R., Boheler, K.R., Taub, D., Hodes, R.J., Longo, D.L., Schlesinger, D., Keller, J., Klotz, E., Kelsoe, G., Umazawa, A., Vesicovi, A.L., Rossant, J., Kunath, T., Hogan, B.L., Curci, A., D'Urso, M., Kelso, J., Hide, W. and Ko, M.S.  
Transcriptome analysis of mouse stem cells and early embryos  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: [cdna@nslg.irc.nia.nih.gov](mailto:cdna@nslg.irc.nia.nih.gov)  
Plate: A0946 Row: A Column: 09  
Seq primer: M13 Reverse  
High quality sequence stop: 315  
POLYA=No.

FEATURES  
source  
1..315  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"







Db	184	TTCTAGGAGAAAACACAGATT	161
RESULT 18			
LOCUS	A0353718	541 bp	DNA
DEFINITION	CITBI-EI-2540U1.TR CITBI-EI Homo sapiens genomic clone 2540U1,		
ACCESSION	A0353718		
VERSION	A0353718.1		
KEYWORDS	GS5.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 541)		
TITLE	Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.		
JOURNAL	Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building		
COMMENT	Unpublished (1997)		
Other GS5s:	CITBI-EI-2540U1.TF		
Contact:	Shaying Zhao, William Nierman, Mark Adams		
The Institute for Genomic Research			
9712 Medical Center Dr., Rockville, MD 20850			
Tel: 301 838 0200			
Fax: 301 838 0208			
Email: hbe@tigr.org			
Clones are available from Research Genetics (info@resgen.com). BAC			
end search page:			
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.			
Seq primer: M13 Reverse			
Class: BAC ends.			
Location/Qualifiers			
1. 541			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/clone="2540U1"			
/sex="male"			
/cell_type="sperm"			
/clone_11b="CITBI-EI"			
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;			
Caltech Human BAC Library D"			
ORIGIN			
Query Match	13.1%;	Score 52.4;	DB 11; Length 541;
Beat Local Similarity	54.7%;	Pred. No. 0.017;	
Matches	98;	Conservative	2; Mismatches 79; Indels 0; Gaps 0;
55	GTGGCGTACATGAGACTGATGCTGTTACTCTCGGAAAGTAGGTATCGACGGGCA	114	
308	GTGGATTCACATACAAAGCTTTGGAGATCTACCCAGATGAGATCGGCTCTCA	367	
115	GTGAGCGCTACCTTTTGTACGTGTTGATATATATATATACATACACACACACACA	174	
368	TATATTTCAACGGCATTTATTTATTTGACACATATATATATATATATACACACACA	427	
175	CACACACATATTTTAACTCGGACATAAATGTACACACACTTATTCAGATTAATACT	233	
428	CACACATATATATATATATATTTTGAACAATGTCACCACTATATTTGAATTAAT	486	
RESULT 19			
LOCUS	A0779979/c	619 bp	DNA
DEFINITION	HS_3113_A1.D12 MR CIT Approved Human Genomic Sperm Library D Homo		
	sapiens genomic clone Plate=3113 Col=23 Row=G, genomic survey		
ACCESSION	A0779979		

VERSION	A0779979.1	GI:5682939
KEYWORDS	GSS.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo	
REFERENCE	1 (bases 1 to 619) Keller,A.,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L. Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)	
TITLE		
JOURNAL	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 3113 row: G column: 23 Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 619.	
FEATURES	Location/Qualifiers	
SOURCE	1..619 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="Plate=3113 Col=23 Row=G" /sex="male" /clone_lib="CIT Approved Human Genomic Sperm Library D" /note="Organ: sperm; Vector: pBelobac11; BAC clones in E-Coli DH10B"	
ORIGIN		
Query Match	13.1%; Score 52.4; DB 11; Length 619;	
Best local similarity	56.5%; Pred. No. 0.017; 72; Indels 0; Gaps 0;	
Matches	95; Conservative 1; Mismatches 72; Indels 0; Gaps 0;	
Db	19 TATATTTACGAGTCTTGAAATAAGCTTAATAATCACTGTGGCGTCACATGAGACTGATGA 78	
Oy	482 TTGATGTTTGATGATGATGAAAAAATTCTCAAGATCTGTGCACAAGAATGATACATGTTAA 423	
Dd	79 TACTGTTTACCTCTGAAAAGTAGGTATGACAGGGCAGAGTGAAGCCCTTCTTTGTACTGT 138	
Oy	422 CACTGCGTACCTGTGCACCTTAAAAATGATGTTGTGACGGTAAATTCGTGTTTTTACCACAAAT 363	
Dd	139 TTGATATATATATATATACATACACACACACACACACACACACACATTA 186	
Oy	362 TAAATATATATATATATATGTCAC 315	
RESULT 20		
AGS61036	736 bp DNA linear GSS 23-DEC-2004	
LOCUS	Mus musculus molossinus DNA, clone:MSWg01-480G02.T7, genomic survey	
DEFINITION	sequence.	
ACCESSION	AGS61036	
VERSION	AGS61036.1 GI:48321734	
KEYWORDS	GSS.	
SOURCE	Mus musculus molossinus (Japanese wild mouse)	
ORGANISM	Mus musculus molossinus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.	
REFERENCE	1 Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saltou,N., Hattori,M., Sakaki,Y., Moriwaiki,K. and	



TITLE	Shiroishi, T.
CONTRIBUTION	Contribution of Asian mouse subspecies <i>Mus musculus molossinus</i> to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL	Genome Res. 14 (12), 2439-2447 (2004)
PUBMED	15574823
REFERENCE	2 (bases 1 to 736)
AUTHORS	Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (17-NOV-2003) Maaahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shiohito-chou, Tsunumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT	Clones are derived from the mouse BAC library MSNg01. For BAC library availability, please contact Kunihisa Abe (abe@tc.riken.jp), The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyada, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@tc.riken.jp
PRIMERS	Sequencing : T7
LIBRARY	Vector : pBAC3.6
R.Site 1	: EcoRI
R.Site 2	: EcoRI
FEATURES	Location/Qualifiers
source	1..736 /organism="Mus musculus molossinus" /mol_type="genomic DNA" /sub_species="molossinus" /db_xref="taxon:57486" /clone="MSNg01-480G02.T7" /sex="male" /tissue_type="mixture of kidney and spleen" /clone_lib="MSNg01 Mouse Male BAC library"
ORIGIN	
Query Match	13.1%; Score 52.4; DB 14; Length 736;
Best Local Similarity	59.7%; Pred. No. 0.017; Indels 0; Gaps 0;
Matches	86; Conservative 1; Mismatches 57;
OR	142 ATATATATATATACATACACACACACACACACACATATTTAACTGGGACA 201
DB	277 AACATACATACACACATACATACACACACACACACACCTATATTTTCA 336
OR	202 TAAATGTACACCACTATTCAAAGTAAACTATCTCAGAGTAAAGCATATCTT 261
DB	337 TATATATATATGATCTACTTAATATGTTAGTGTCATTTATATCAGAAAAATATCAT 396
OR	262 GCTATTGTCATTAACACTAGCAGTTA 285
DB	397 CCTAATACCTATGCTTCATTA 420
RESULT 21	
AKI42564/c	AKI42564
LOCUS	2558 bp mRNA linear HTC 21-SEP-2005
DEFINITION	Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630038012 product:unclassified, full insert sequence.
ACCESSION	AKI42564
VERSION	AKI42564.1 GI:74216314
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE	1 Carninci, P. and Hayashizaki, Y.
AUTHORS	High-efficiency full-length cDNA cloning
TITLE	

JOURNAL PUBMED REFERENCES	Mech. Enzymol. 303, 19-44 (1999)
AUTHORS	2
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komou, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL PUBMED REFERENCE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	3
Kono, H., Akiyama, J., Nishi, K., Kitamura, T., Tashtiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)	
11076861	4
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Atawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Salto, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Salto, R., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T., Fleischmann, U., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kusih, P., Lewis, S., Matsuo, Y., Nikaido, I., Perole, G., Quackenbush, J., Schriml, L.M., Seabill, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, D., Boftelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochman, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Machioni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, R., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Togo-Oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L., Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.	
Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
11217851	5
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Mikado, I., Osato, N., Salto, R., Suzuki, H., Yamana, A., Kiyosawa, H., Yagi, K., Tomaru, Y., Haegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanpin, A., Matsuda, H., Batalova, S., Beisel, K.W., Blake, A., Brad, D., Brusch, V., Chotaha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragan, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawasaki, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kuochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nmatas, K., Okido, T., Pavan, W.M., Pettes, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Rindang, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultan, A., Takemura, Y., Taylor, M.S., Teasdale, R.D., Tomita, K., Veardo, R., Wagner, L., Wanless, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hironaka-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasuniishi, A., Yoshino, M., Waterston, R., Lander, E.S.,	











DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 74-144, >(TA)n#simple\_repeat (matched complement)  
 157-394, >LIM3A#LINE/L1 157-333, >L1MB7#LINE/L1  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

Location/Qualifiers  
 1..665  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-B-EJ1-a18-h-14-0-UI"  
 /issue\_type="fetal eyes, lens, eye anterior segment,  
 optic nerve, retina, Retina Foveal and Macular, RPE and  
 Choroid"  
 /dev\_stage="fetal and adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-B-EJ1"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Ecor I; Site\_2: Not I;  
 UI-B-EJ1 is a subcloned cDNA library constructed  
 according to Bonaldo, Lennon and Soares, Genome Research,  
 6:791-806, 1996. First strand cDNA synthesis was primed  
 with an oligo-dT primer containing a Not I site. Double  
 stranded cDNA was ligated to an Ecor I adaptor, digested  
 with Not I, and cloned directionally into pT73-Pac  
 vector. The oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (dT)18 tail. The  
 sequence tags for this library are: fetal eyes,  
 AGAATCAGCA; lens, CGATTAGCGA; eye anterior segment,  
 AAGCGCGCAT; optic nerve, CCATTAGTC; retina, CCGCG; Retina  
 Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This  
 library was created for the program, Gene Discovery in the  
 Visual System, supported by National Eye Institute (NEI).  
 TAG TISSUE=Foveal and Macular Retina  
 TAG LIB=UI-B-EJ1  
 TAG\_SEQ=GTCC"

## ORIGIN

Query Match 13.1%; Score 52.2; DB 3; Length 665;  
 Best Local Similarity 72.5%; Pred. No. 0.019;  
 Matches 66; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

115 GTGAGGCTTACCTTTTGTACTGTTGATATATATATATACACACACACACA 174  
 |||||  
 67 GGGAGGATATATATATATATATATATATATATATATATACACACACACA 126  
 |||||  
 175 CACACACATATTTTAACTGGGACATATA 205  
 |||||  
 127 CACACACATATATATATATGTCGTCTATAGA 157  
 |||||

RESULT 25  
 DW264376 820 bp mRNA linear EST 05-JAN-2006  
 LOCUS DW264376  
 DEFINITION UI-S-GN1-abn-d-03-0-UI.61 UI-S-GN1 Euprymna scolopes cDNA clone  
 ACCESSION DW264376  
 VERSION DW264376.1 GI:84429779  
 KEYWORDS EST.

## SOURCE

ORGANISM Euprymna scolopes  
 Euprymna scolopes  
 Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;  
 Decapodiformes; Sepioidae; Sepioidae; Euprymna.

## REFERENCE

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 PUBMED 8889548

## COMMENT

Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. M. J. McFall-Ngai, University of  
 Wisconsin-Madison  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/squid.html>  
 The following repetitive elements were found in this cDNA  
 sequence: 306-354, >(CA)n#simple\_repeat (matched complement)  
 518-546, >AT-rich#low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

Location/Qualifiers  
 1..820  
 /organism="Euprymna scolopes"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6613"  
 /clone="UI-S-GN1-abn-d-03-0-UI"  
 /issue\_type="Euprymna scolopes 12-hour symbiotic"  
 /dev\_stage="12 hours symbiotic"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-S-GN1"  
 /note="Vector: pT73-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Ecor I; Site\_2: Not I; UI-S-GN1 is a  
 normalized library derived from UI-S-GN1. The library was  
 constructed and normalized according to Bonaldo, Lennon  
 and Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an Ecor I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this  
 library is GAGCCTGACT  
 TAG TISSUE=Euprymna\_scolopes\_12-hour\_symbiotic  
 TAG LIB=UI-S-GN1  
 TAG\_SEQ=GAGCCTGACT"

## ORIGIN

Query Match 13.1%; Score 52.2; DB 10; Length 820;  
 Best Local Similarity 65.2%; Pred. No. 0.019;  
 Matches 75; Conservative 1; Mismatches 39; Indels 0; Gaps 0;

96 AAGTAGGTATGACAGGAGGAGGCTTCTTTGTACTGTTGATATATATATAC 155  
 |||||  
 396 ATGGGCTATCTTTTGTATATATTTCTTTCTTTTATATATATATATATAC 337  
 |||||  
 156 ATACACACACACACACACACATATTTTAACTGGGACATATAATGTAC 210  
 |||||  
 336 ACACACACACACACACACACATTTTGAAGGCAAAAAAGCTTC 282  
 |||||

RESULT 26  
 DU046739 950 bp DNA linear GSS 12-AUG-2005  
 LOCUS DU046739  
 DEFINITION clone LB\_HB80036P17 5, genomic survey sequence.  
 ACCESSION DU046739  
 VERSION DU046739.1 GI:72473946  
 KEYWORDS GSS.

## SOURCE

ORGANISM Lycopersicon esculentum (Solanum lycopersicum)  
 Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.



Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: GGAAACGCTATTGACATG

REFERENCE  
AUTHORS  
1 (bases 1 to 799)  
Adams, D. J., Biggs, P. J., Cox, A. V., Davies, R. M., van der Weyden, L.,  
Jonkers, J., Smith, J., Plumb, R. W., Taylor, R. G., Nishijima, I., Yu, Y.,  
Rogers, J. and Bradley, A.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,











	Query Match	Best Local Similarity	Matches	Score	No. DB	Length	Gaps
	13.0%	58.9%	89	51.8	DB 13	669	0
				Pred. No. 0.024		Indels	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches	62
						Conservative	0
						Matches	89
						Indels	0
						Gaps	0
						Mismatches</	

[illegible]



```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-27539"
/tissue_type="Testis"
/note="vector pindigBAC-536"

```

Query Match	Best Local Similarity	12.9%	Score 51.6	DB 14	Length 827	
Matches	81	Conservative	1	Mismatches 50	Indels 0	Gaps 0
QY	142	ATATATATATATACATACACACACACACACACACATATTTAACTCGGAC	201			
DB	816	ACACACACACACACACACACACACACACACACAAATTATAGCAACATTA	757			
QY	202	TAAATGACACCACTATTCAAAGTAAACCTATCTCGAAAGTAAGCATTAATCCT	261			
DB	756	CAATTATTTCTTTTCAAAATGAAGTCACTGTATTTTGAGAGTTCATTGGAATCCA	697			
QY	262	GCTATGTGCATT	273			
DB	696	TCTAATAGCATT	685			

RESULT 36	BH43871/c	528 bp	DNA	linear	GSS 03-DEC-2001
LOCUS	BH43871				
DEFINITION	CH230-41IN10.TV CHORI-230 Segment 1 <i>Rattus norvegicus</i> genomic clone				
DESCRIPTION	CH230-41IN10, genomic survey sequence.				

accession  
 BH343871  
 version  
 BH343871.1  
 keywords  
 GSS.  
 Rattus norvegicus (Norway rat)  
 source

REFERENCE  
1 (bases 1 to 528)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Rattus.

REFERENCE	1 (bases 1 to 528)
AUTHORS	Zhao, S., Shetty, J., Shatman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregorjisi, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.
TITLE	Rat BAC End sequences from Library CHORI-230 EcoRI segment
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: CH230-41N10..TJ

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhac@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pierler de Jong ([pdjong@mail.cho.org](mailto:pdjong@mail.cho.org)).  
Clones may be purchased from BACPAC Resources  
(<http://www.chori.org/bacpac/ordering/information.htm>). BAC end  
page: [http://www.tigr.org/cdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html)  
Plate: 41 row: N column: 10  
Seq primer: T7  
Class: BAC ends.

```

FEATURES
    source
        location/Qualifiers
            1..528
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /strain="BN/SanHsd/MCw"
                /db_xref="taxon:10116"
                /clone="CH230-41N10"
                /sex="Female"
                /cell_type="Brain"
                /clone_1ib="CHORI-230 Segment 1"
                /note="Vector: pINR6AC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SanHsd/MCw) BAC library produced by

```

ORIGIN	Pieter de Jong
Query Match	12.9%; Score 51.4; DB 11; Length 528;
Best Local Similarity	61.2%; Pred. No. 0.03; Indels 0;
Matches 79; Conservative	2; Mismatches 48; Gaps 0;

QY 58 GCGTACCATGACTGCTGATTAAGTGGTACCTGTGAAAGTAGGTATGCAAGGCAGATG 117

Db 497 GAGAGATATGAGGCTGTTAATTTGGGATATGCTCTATTTCAGAGAGATGTGATAGC 438

QY 118 AGGCTTACTTTTGTACTGTTGATATATATATACATACACACACACACAC 177

Db 437 GAGCATATATATATATATATATATATATATATATACACACACACACACACAC 378

QY 178 ACAACATTA 186

Db 377 ACAACACAC 369

RESULT	37
BZ250267	
LOCUS	BZ250267
DEFINITION	BZ250267 777 bp DNA linear GSS 12-OCT-2007
CH230-353M6.TV CHOR1-230 Segment 2 Rattus norvegicus genomic clone	
ACCSSION	CH230-353M6, genomic survey sequence.
VERSION	BZ250267
KEYWORDS	BZ250267.1 GI:23911520
SOURCE	GSS.
ORGANISM	Rattus norvegicus (Norway rat)
	Rattus norvegicus

REFERENCE  
AUTHORS

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,  
1 (bases 1 to 777)  
Eumariota; Metazoa; Chordata; Crinata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Scurionathi; Muridae; Murinae; Rattus.

TITLE Rat BAC End Sequences from Library CHOR1-230 Mbol segment  
 JOURNAL Unpublished (1989)  
 COMMENT Other\_GSSs: CH230-353M6.TU  
 Other\_GSSs:

Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel.: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat330.htm>). For BAC library  
availability, please contact Pierer de Jong ([pdjong@small.cho.org](mailto:pdjong@small.cho.org))  
Clones may be purchased from BACDAC Resources  
([http://www.chori.org/bacpac/orering\\_information.htm](http://www.chori.org/bacpac/orering_information.htm)). BAC end  
page: [http://www.tigr.org/cdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html)  
Plate: 353 row: M column: 6  
Seq primer: T7  
Class: BAC ends.

```

FEATURES
source
Location/Qualifiers
1..777
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SaHsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-353M6"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHOR1-230 Segment 2"
/notes="Vector: pTARAC1.3; Site_1: MboI;
CHOR1-230 Rat (BN/SaHsd/MCw) BAC library produced by
Pieter de Jong"
ORIGIN
Query Match      12.9%; Score 51.4; DB 11; Length 777;
Best Local Similarity 60.7%; Pred. No. 0.031;

```



	Matches	82;	Conservative	1;	Mismatches	52;	Indels	0;	Gaps	0;
OY	136	TGTTTGATATATATATATACATACACACACACACACACACACATATTTTAATCTC	195							
Db	191	TGATTTTAAAAATCATCACACACACACACACACACACACACACACACACACAC	250							
OY	196	GGGACATTAATGTATACACCACTATTCCAAGTAAAACTATCTCAGAGATTAAGACAT	255							
Db	251	CACACTGATGGTTGTGCTTAACCTTAGCAGCACACATATCTCTTCCAAATTAAGAAT	310							
OY	256	AATCTGCTATTTGTC	270							
Db	311	AGTATTCAATTGGC	325							

RESULT	38
AZ514487	
LOCUS	
DEFINITION	AZ514487 178 bp DNA linear GSS 05-OCT-2001
ACCESSION	U00361F24 Mouse 10kb plasmid UUC1M library Mus musculus genomic
VERSION	clone UUC1M0361F24 F, genomic survey sequence.
KEYWORDS	AZ514487 AZ514487 GI:10695899
SOURCE	GSS.
ORGANISM	Mus musculus (house mouse)
	Mus musculus

REFERENCE  
AUTHORS  
1 (Pages 1 to 178)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
et al. "Method of Measuring the Effect of the

TITLE	Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL	plasmid inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0361 row: F column: 24  
Seq primer: CGTTGTAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 178.

```

FEATURES
Source
Location/Qualifiers
1..178
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0361F24"
/sex="Male"
/lab_host="B. Coli strain XL10-GOLD, Tl-resistant, F-"
/clone.lib="Mouse 10kb plasmid uucg1m library"
/notes="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated

```

with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	12.8%;	Score 51.2;	DB 11;	Length 178;
Best Local Similarity	63.1%;	Pred. No. 0.031;		
Matches 77; Conservative	1;	Mismatches 44;	Indels 0;	Gaps 0;

QY 202 TAAATGTACACCACTATTCCAAAGTAAAACTATCTCAGAAAGTAAAGCATATATCTT 261

Db 63 CACCTATTAATATACGACATATATATATATATACGACATATCATTAAGTCTCTCAAGCAT 122

QY 262 GC 263

Db 123 GC 124

RESULT 39	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AZ703674/c	AZ703674	541 bp	DNA	linear	GSS	24-JAN-2001	
	RPCT-23-226623.TV	Mus musculus genomic clone					
	RPCT-23-226623	genomic survey sequence.					
	AZ703674						
	AZ703674.1	GI:12428297					
	GSS.						
	Mus musculus	(house mouse)					
	Mus musculus						

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 541)  
Zhao, S., Nierman, W., Feldbljum, T., Malek, D., Shatsman, S.,  
Alinest, B., Levins, M., McGann, S., Teegye, G., Geer, K., Krol, M., de  
Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RRC1-23  
Unpublished (1999)  
Contact: Shaying Zhao

```

Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPC1-23. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 226 Row: G Column: 23
Seq primer: T7
Class: BAC ends
Location/Qualifiers
1..541
source

```

```

FEATURES
    source
        location/Qualifiers
            1..541
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPIC1-23-226G23"
                /sex="Female"
                /lab_host="DH10B"
                /clone_idb="RPIC1-23"
                /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1:
                EcorI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or
                brain genomic DNA was isolated and partially digested
                with a combination of EcorI and EcorI methylase. Site
                selected DNA was cloned into the pBACe3.6 vector at the

```



ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). \*

KEYWORDS EST.  
SOURCE Prunus persica (peach)  
ORGANISM Prunus persica

Query Match 12.8%; Score 51.2; DB 11; Length 541;  
Best Local Similarity 60.1%; Pred. No. 0.034;  
Matches 83; Conservative 1; Mismatches 54; Indels 0; Gaps 0;

REFERENCE  
AUTHORS Abbott, A., Zhebentyayeva, T., Main, D., Jung, S., Staton, M.,  
TITLES Uesudurai, C. and Wing, R.  
JOURNAL Peach Model Genome for Rosaceae Shoot ESTs  
COMMENT Unpublished (2006)  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 601  
Seq primer: TAAATGACATCTACTATAGGG  
High quality sequence stop: 607.  
Location/Qualifiers

RESULT 40  
CT454907 750 bp DNA linear GSS 03-NOV-2005  
LOCUS Sus scrofa genomic clone CH242-455A2, genomic survey sequence.  
ACCESSION CT454907  
VERSION CT454907.1 GI:80174047  
KEYWORDS GSS.  
SOURCE Sus scrofa  
ORGANISM Sus scrofa (pig)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
Sus.  
1 (bases 1 to 750)  
Humphray, S.J., Plumb, R.W. and Durham, J.L.  
Direct Submission  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humphray@sanger.ac.uk Unpublished  
This sequence was generated from the T7 end of BAC 455A2. 455A2 is  
part of the CHORI-242 BAC library created by P. de Jong. Further  
details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/).  
Location/Qualifiers

FEATURES  
source  
1. 607  
/organism="Prunus persica"  
/mol\_type="mRNA"  
/cultivar="Loring"  
/db\_xref="taxon:3760"  
/clone="PP\_Lec0001P09f"  
/rname\_type="Shoot"  
/lab\_host="E. coli"  
/clone\_1lb="Peach shoot"  
/note="Vector: pBluescript II SK(-); Site\_1: EcorI;  
Site\_2: XhoI"

REFERENCE  
AUTHORS Humphray, S.J., Plumb, R.W. and Durham, J.L.  
TITLES Direct Submission  
JOURNAL Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humphray@sanger.ac.uk Unpublished  
This sequence was generated from the T7 end of BAC 455A2. 455A2 is  
part of the CHORI-242 BAC library created by P. de Jong. Further  
details: [http://www.sanger.ac.uk/Projects/S\\_scrofa/](http://www.sanger.ac.uk/Projects/S_scrofa/).  
Location/Qualifiers

1. 750  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-455A2"  
/rname\_type="white blood cells"  
/note="vector pTRABAC1.3\_BamHI  
sex female"

ORIGIN

Query Match 12.8%; Score 51.2; DB 14; Length 750;  
Best Local Similarity 69.4%; Pred. No. 0.035;  
Matches 66; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

RESULT 42  
A0326274/c 607 bp DNA linear GSS 08-JAN-1999  
LOCUS nxb0028m24f CGCI Rice BAC Library Oryza sativa (japonica  
DEFINITION cultivar-group) genomic clone nxb0028m24f, genomic survey  
sequence.  
ACCESSION A0326274  
VERSION A0326274.1 GI:4118124  
KEYWORDS GSS.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicotyledons;  
rosids; eurosid I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 607)  
Zhebentyayeva, T., Main, D., Jung, S., Staton, M.,  
Uesudurai, C. and Wing, R.  
Peach Model Genome for Rosaceae Shoot ESTs  
Unpublished (2006)  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 601  
Seq primer: TAAATGACATCTACTATAGGG  
High quality sequence stop: 607.  
Location/Qualifiers

FEATURES  
source  
1. 750  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-455A2"  
/rname\_type="white blood cells"  
/note="vector pTRABAC1.3\_BamHI  
sex female"

FEATURES  
source  
1. 607  
/organism="Prunus persica"  
/mol\_type="mRNA"  
/cultivar="Loring"  
/db\_xref="taxon:3760"  
/clone="PP\_Lec0001P09f"  
/rname\_type="Shoot"  
/lab\_host="E. coli"  
/clone\_1lb="Peach shoot"  
/note="Vector: pBluescript II SK(-); Site\_1: EcorI;  
Site\_2: XhoI"

ORIGIN

Query Match 12.8%; Score 51.2; DB 14; Length 750;  
Best Local Similarity 69.4%; Pred. No. 0.035;  
Matches 66; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

RESULT 42  
A0326274/c 607 bp DNA linear GSS 08-JAN-1999  
LOCUS nxb0028m24f CGCI Rice BAC Library Oryza sativa (japonica  
DEFINITION cultivar-group) genomic clone nxb0028m24f, genomic survey  
sequence.  
ACCESSION A0326274  
VERSION A0326274.1 GI:4118124  
KEYWORDS GSS.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicotyledons;  
rosids; eurosid I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 607)  
Zhebentyayeva, T., Main, D., Jung, S., Staton, M.,  
Uesudurai, C. and Wing, R.  
Peach Model Genome for Rosaceae Shoot ESTs  
Unpublished (2006)  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 601  
Seq primer: TAAATGACATCTACTATAGGG  
High quality sequence stop: 607.  
Location/Qualifiers

Query Match 12.8%; Score 51.2; DB 14; Length 750;  
Best Local Similarity 69.4%; Pred. No. 0.035;  
Matches 66; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

RESULT 42  
A0326274/c 607 bp DNA linear GSS 08-JAN-1999  
LOCUS nxb0028m24f CGCI Rice BAC Library Oryza sativa (japonica  
DEFINITION cultivar-group) genomic clone nxb0028m24f, genomic survey  
sequence.  
ACCESSION A0326274  
VERSION A0326274.1 GI:4118124  
KEYWORDS GSS.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicotyledons;  
rosids; eurosid I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 607)  
Zhebentyayeva, T., Main, D., Jung, S., Staton, M.,  
Uesudurai, C. and Wing, R.  
Peach Model Genome for Rosaceae Shoot ESTs  
Unpublished (2006)  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 601  
Seq primer: TAAATGACATCTACTATAGGG  
High quality sequence stop: 607.  
Location/Qualifiers

RESULT 41  
DM340951 607 bp mRNA linear EST 06-JAN-2006  
LOCUS PP\_Lec0001P09f Peach shoot Prunus persica cDNA clone  
DEFINITION PP\_Lec0001P09f, mRNA sequence.  
ACCESSION DM340951  
VERSION DM340951.1 GI:84569332

REFERENCE  
AUTHORS Wing, R.A. and Dean, R.A.  
TITLES A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University







Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szabo@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 299 row: K column: 23  
Seq primer: T7

Class: BAC ends.

# FEATURES

## source

1. .672  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-299K23"  
/sex="Female"  
/lab\_host="DH10B"  
/clone\_lib="RPCI-23"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

## ORIGIN

Query Match 12.8%; Score 51; DB 11; Length 672;  
Best Local Similarity 59.6%; Pred. No. 0.038;  
Matches 84; Conservative 1; Mismatches 56; Indels 0; Gaps 0;  
QY 78 ATATCTGTTACCTCTGAAAAGTAGATGCGAGGCACTGACCTTTTGTACTG 137  
DB 532 ATATGTGAGACCCCTCTCAAAAATAGACATTAGAAATATCATCTATTATATATTC 591  
QY 138 TTGAT 197  
DB 592 ATGAT 651  
QY 198 GACATTAATGTACACCACTTA 218  
DB 652 TAAAACTGCTACTAGAGCTA 672

RESULT 45  
CT398920/c 835 bp DNA linear GSS 03-NOV-2005  
LOCUS  
DEFINITION  
Sus scrofa genomic clone CH242-402G15, genomic survey sequence.  
ACCESSION  
CT398920.1 GI:80050005  
VERSION  
CT398920.1 GI:80050005  
KEYWORDS  
GSS.  
SOURCE  
Sus scrofa (pig)  
ORGANISM  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
Sus.

REFERENCE  
1 (bases 1 to 835)  
Humphray, S.J., Plumb, R.W. and Durham, J.L.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (01-NOV-2005) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished

COMMENT  
This sequence was generated from the SPE end of BAC 402G15. 402G15 is part of the CHOBI-242 BAC library created by P. de Jong. Further details: http://www.sanger.ac.uk/Projects/S\_scrofa/.

## FEATURES

Location/Qualifiers

## source

1. .835  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="CH242-402G15"  
/tissue\_type="white blood cells"  
/note="vector pTRBAC1.3\_BamHI  
sex female"

## ORIGIN

Query Match 12.8%; Score 51; DB 14; Length 835;  
Best Local Similarity 59.6%; Pred. No. 0.039;  
Matches 84; Conservative 1; Mismatches 56; Indels 0; Gaps 0;  
QY 97 AGTAGGTATGCGAGGCAAGTAGCCCTTCTTTGTAAGTGTATATATATATCA 156  
DB 757 ACTGTGATGAGACATGATGAGAGAGATGAGTAATGTGTATATATATACACA 698  
QY 157 TACACACACACACACACACACACACACATATATATATATATATATATATATACACACC 216  
DB 697 CACACACACACACACACACACACACATATATATATATATATATATATATATATATAT 638  
QY 217 TATTCAAAGTAAACTACTA 237  
DB 637 AATTGACATCAATTATAGTA 617

## RESULT 46

DM594541 1072 bp mRNA linear EST 18-JAN-2006  
LOCUS  
DEFINITION  
CGX113-F04 5', mRNA sequence.

ACCESSION  
DM594541  
VERSION  
DM594541.1 GI:85167999  
KEYWORDS  
EST.  
SOURCE  
Gasterosteus aculeatus (three spined stickleback)  
ORGANISM  
Gasterosteus aculeatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
Gasterosteidae; Gasterosteus.  
1 (bases 1 to 1072)

REFERENCE  
Kingstley, D.M., Peichel, C., Knecht, A., Balabhadra, S., Grimwood, J.,  
Dickson, M., Schmutz, J., and Myers, R.W.  
Expressed sequence tags from Gasterosteus aculeatus (2004)  
TITLE  
Unpublished (2004)  
JOURNAL  
Contact: Grimwood, Jane  
Stanford Human Genome Center  
Stanford University School of Medicine  
975 S California Ave, Palo Alto, CA 94304, USA  
Tel: 650 320 5917  
Fax: 650 320 5801  
Email: jane@hgsc.stanford.edu  
Plate: 113

## FEATURES

## source

High quality sequence stop: 855.  
Location/Qualifiers

1. .1072  
/organism="Gasterosteus aculeatus"  
/mol\_type="mRNA"  
/strain="Conner Creek sticklebacks, WA USA"  
/db\_xref="taxon:69293"  
/clone="CGX113-F04"  
/sex="mixed male and female"  
/tissue\_type="eyes"  
/dev\_stage="adult"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer:







JOURNAL PUBMED REFERENCE AUTHORS	Genome Res. 10 (11), 1757-1771 (2000) 11076861 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Iwama, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flisbachmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabli, F., Suzuki, R., Tomita, M., Wanger, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyono, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Borja, A., Yoshida, K., Haesegawa, Y., Kawaji, H., Kontsuki, S. and Hayashizaki, Y.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 11217851 5 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oseko, N., Saito, R., Suzuki, H., Yamanka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Haesegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chachua, C., Corbani, L.E., Cousins, S., Dalla, B., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lenhard, R., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numa, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B., Ringwald, M., Sanderlin, A., Schneider, C., Sempole, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Borja, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shihama, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420 (6915), 563-573 (2002) 12466851 6 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodlun, R., Shimokawa, K., Bajic, V.B., Bremner, S.B., Batalov, S., Foxe, A.R., Zavesan, M., Davis, M.J., Wilming, L.G., Adkins, V., Allen, J.E., Ambesi-Imbondato, A., Apweiler, R., Aurilio, R., N., Bailey, T.L., Banerji, M., Baxter, L., Beisel, K.W., Berzano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crome, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N.,
JOURNAL PUBMED REFERENCE AUTHORS	Hill, D., Humnicki, L., Iacono, M., Ikeo, K., Iwama, A., Iehikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelsio, J., Kitamura, H., Kitano, H., Kollas, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Lium, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakachi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, P., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavoni, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.P., Ring, B.Z., Ringwald, M., Rott, B., Ruan, Y., Salzberg, S.L., Sanderlin, A., Schneider, C., Schonbach, C., Sekizuki, K., Sempole, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugita, K., Sultana, R., Takenaka, Y., Taki, K., Tammo, K., Tan, S.L., Tang, S., Taylor, M.S., Tegen, J., Teichmann, S.A., Ueda, H., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamaniishi, H., Zaslavsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlstedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori, Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Komno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessey, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watanishi, A., Okamura-Oho, Y., Suzuki, H., Kawai, J. and Hayashizaki, Y.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Functional annotation of the mammalian genome Science 309 (5740), 1559-1563 (2005) 16141072 7 Karyama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yag, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sanderlin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlstedt, C.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	RIKEN Genome Exploration Research Group Antisense transcription in the mammalian transcriptome Science 309 (5740), 1564-1566 (2005) 16141073 8 (bases 1 to 1722) Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hori, P., Iida, J., Imamura, K., Imotani, K., Itoh, M., Kanagawa, S., Kawai, J., Kojima, M., Komno, H., Murata, M., Nakamura, M., Ninomiya, N., Nishiyori, H., Nomura, K., Ohno, M., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shiraki, T., Tagami, M., Tagami, Y., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Direct Submission Submitted (30-MAR-2004) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 220-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ Location/Qualifiers 1. 1722 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB020004018" /db_xref="taxon:10090" /clone="B020004018" /tissue_type="egg"
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420 (6915), 563-573 (2002) 12466851 6 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodlun, R., Shimokawa, K., Bajic, V.B., Bremner, S.B., Batalov, S., Foxe, A.R., Zavesan, M., Davis, M.J., Wilming, L.G., Adkins, V., Allen, J.E., Ambesi-Imbondato, A., Apweiler, R., Aurilio, R., N., Bailey, T.L., Banerji, M., Baxter, L., Beisel, K.W., Berzano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crome, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N.,
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 11217851 5 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oseko, N., Saito, R., Suzuki, H., Yamanka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Haesegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chachua, C., Corbani, L.E., Cousins, S., Dalla, B., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lenhard, R., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numa, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B., Ringwald, M., Sanderlin, A., Schneider, C., Sempole, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Borja, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shihama, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
CONSRMT TITLE JOURNAL PUBMED REFERENCE AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420 (6915), 563-573 (2002) 12466851 6 Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodlun, R., Shimokawa, K., Bajic, V.B., Bremner, S.B., Batalov, S., Foxe, A.R., Zavesan, M., Davis, M.J., Wilming, L.G., Adkins, V., Allen, J.E., Ambesi-Imbondato, A., Apweiler, R., Aurilio, R., N., Bailey, T.L., Banerji, M., Baxter, L., Beisel, K.W., Berzano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crome, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N.,







GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 17, 2006, 17:09:12 ; Search time 151 Seconds  
(without alignments)  
4956.580 Million cell updates/sec

Title: SEQ1-33670G  
Perfect score: 399.2  
Sequence: 1 aagatgctctctctctata.....taactcagatcagacgag 400

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

1: /EMC\_Celerra\_SIDS3/prodata/2/ina/1 COMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/prodata/2/ina/5 COMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/prodata/2/ina/6A COMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/prodata/2/ina/6B COMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/prodata/2/ina/7 COMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/prodata/2/ina/H COMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/prodata/2/ina/PC/US COMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/prodata/2/ina/PP COMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/prodata/2/ina/RE COMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	52.4	13.1	601	US-09-949-016-122917	Sequence 122917, A
2	52.4	13.1	601	US-09-949-016-123011	Sequence 123011, A
3	52.4	13.1	601	US-09-949-016-123105	Sequence 123105, A
4	52.4	13.1	601	US-09-949-016-123199	Sequence 123199, A
5	52.4	13.1	601	US-09-949-016-123293	Sequence 123293, A
6	52.4	13.1	3455	US-10-009-332-31	Sequence 31, Appl
7	51.2	12.8	601	US-09-949-016-122919	Sequence 122919, A
8	51.2	12.8	601	US-09-949-016-122922	Sequence 122922, A
9	51.2	12.8	601	US-09-949-016-122922	Sequence 122922, A
10	51.2	12.8	601	US-09-949-016-123013	Sequence 123013, A
11	51.2	12.8	601	US-09-949-016-123016	Sequence 123016, A
12	51.2	12.8	601	US-09-949-016-123107	Sequence 123107, A
13	51.2	12.8	601	US-09-949-016-123108	Sequence 123108, A
14	51.2	12.8	601	US-09-949-016-123110	Sequence 123110, A
15	51.2	12.8	601	US-09-949-016-123201	Sequence 123201, A
16	51.2	12.8	601	US-09-949-016-123202	Sequence 123202, A
17	51.2	12.8	601	US-09-949-016-123204	Sequence 123204, A
18	51.2	12.8	601	US-09-949-016-123205	Sequence 123205, A
19	51.2	12.8	601	US-09-949-016-123295	Sequence 123295, A
20	51.2	12.8	601	US-09-949-016-123296	Sequence 123296, A
21	51.2	12.8	601	US-09-949-016-123298	Sequence 123298, A
22	51.2	12.8	64377	US-09-949-016-15212	Sequence 15212, A
23	51.2	12.8	64377	US-09-949-016-15213	Sequence 15213, A

24	51.2	12.8	64377	US-09-949-016-15214	Sequence 15214, A
25	51.2	12.8	64377	US-09-949-016-15215	Sequence 15215, A
26	51.2	12.8	64377	US-09-949-016-15216	Sequence 15216, A
27	51	12.8	3462	US-10-009-332-30	Sequence 30, Appl
28	51	12.8	3464	US-10-009-332-26	Sequence 26, Appl
29	51	12.8	3467	US-10-009-332-25	Sequence 25, Appl
30	51	12.8	3467	US-10-009-332-29	Sequence 29, Appl
31	51	12.8	3469	US-10-009-332-27	Sequence 27, Appl
32	51	12.8	3470	US-10-009-332-28	Sequence 28, Appl
33	51	12.8	3473	US-10-009-332-24	Sequence 24, Appl
34	50.8	12.7	601	US-09-949-016-122915	Sequence 122915, A
35	50.8	12.7	601	US-09-949-016-122916	Sequence 122916, A
36	50.8	12.7	601	US-09-949-016-123009	Sequence 123009, A
37	50.8	12.7	601	US-09-949-016-123010	Sequence 123010, A
38	50.8	12.7	601	US-09-949-016-123103	Sequence 123103, A
39	50.8	12.7	601	US-09-949-016-123104	Sequence 123104, A
40	50.8	12.7	601	US-09-949-016-123197	Sequence 123197, A
41	50.8	12.7	601	US-09-949-016-123198	Sequence 123198, A
42	50.8	12.7	601	US-09-949-016-123291	Sequence 123291, A
43	50.8	12.7	601	US-09-949-016-123292	Sequence 123292, A
44	50	12.5	601	US-09-949-016-122921	Sequence 122921, A
45	50	12.5	601	US-09-949-016-123015	Sequence 123015, A
46	50	12.5	601	US-09-949-016-123109	Sequence 123109, A
47	50	12.5	601	US-09-949-016-123203	Sequence 123203, A
48	50	12.5	601	US-09-949-016-123297	Sequence 123297, A
49	49.8	12.5	2534	US-09-526-193A-23	Sequence 23, Appl
50	49.8	12.5	153866	US-09-949-016-16919	Sequence 16919, A
51	49.6	12.4	601	US-09-949-016-62380	Sequence 62380, A
52	49.6	12.4	601	US-09-949-016-122918	Sequence 122918, A
53	49.6	12.4	601	US-09-949-016-123012	Sequence 123012, A
54	49.6	12.4	601	US-09-949-016-123106	Sequence 123106, A
55	49.6	12.4	601	US-09-949-016-123200	Sequence 123200, A
56	49.6	12.4	601	US-09-949-016-123294	Sequence 123294, A
57	49.6	12.4	601	US-09-949-016-187550	Sequence 187550, A
58	49.4	12.4	312470	US-09-949-016-14043	Sequence 14043, A
59	49.4	12.4	336024	US-09-949-016-12373	Sequence 12373, A
60	48.6	12.2	601	US-09-949-016-87162	Sequence 87162, A
61	48.6	12.2	140840	US-09-949-016-14199	Sequence 14199, A
62	48.6	12.2	786431	US-09-751-389-3	Sequence 3, Appl
63	48.4	12.1	601	US-09-949-016-62379	Sequence 62379, A
64	48.4	12.1	66931	US-09-949-016-187549	Sequence 187549, A
65	48.4	12.1	66933	US-09-544-398B-11	Sequence 11, Appl
66	48.4	12.1	66933	US-09-543-771B-11	Sequence 11, Appl
67	48.4	12.1	72049	US-09-543-398B-9	Sequence 9, Appl
68	48.4	12.1	72049	US-09-543-771B-9	Sequence 9, Appl
69	48.4	12.1	98962	US-09-949-016-14133	Sequence 14133, A
70	48.4	12.1	102884	US-09-949-016-17100	Sequence 17100, A
71	48.2	12.1	134890	US-09-949-016-15602	Sequence 15602, A
72	48.2	12.1	183202	US-09-949-016-13614	Sequence 13614, A
73	48	12.0	601	US-09-949-016-56477	Sequence 56477, A
74	48	12.0	601	US-09-949-016-56478	Sequence 56478, A
75	48	12.0	601	US-09-949-016-56479	Sequence 56479, A
76	48	12.0	601	US-09-949-016-113290	Sequence 113290, A
77	48	12.0	601	US-09-949-016-159266	Sequence 159266, A
78	48	12.0	148794	US-09-949-016-122751	Sequence 122751, A
79	48	12.0	285986	US-09-949-016-12287	Sequence 12287, A
80	48	12.0	288031	US-09-949-016-14864	Sequence 14864, A
81	48	12.0	636591	US-09-949-016-113808	Sequence 113808, A
82	48	12.0	636591	US-09-949-016-133388	Sequence 133388, A
83	47.8	12.0	601	US-09-949-016-82382	Sequence 82382, A
84	47.8	12.0	601	US-09-949-016-187552	Sequence 187552, A
85	47.8	12.0	93778	US-09-949-016-15096	Sequence 15096, A
86	47.8	12.0	209631	US-09-949-002-574	Sequence 574, App
87	47.8	12.0	209632	US-09-949-002-802	Sequence 802, App
88	47.6	11.9	601	US-09-949-016-174330	Sequence 174330, A
89	47.6	11.9	101349	US-09-949-016-12762	Sequence 12762, A
90	47.6	11.9	117937	US-09-949-016-15775	Sequence 15775, A
91	47.6	11.9	117937	US-09-949-016-15775	Sequence 15775, A
92	47.6	11.9	122626	US-09-949-016-17524	Sequence 17524, A
93	47.4	11.9	601	US-09-949-016-112707	Sequence 112707, A
94	47.4	11.9	601	US-09-949-016-112709	Sequence 112709, A
95	47.4	11.9	601	US-09-949-016-112791	Sequence 112791, A
96	47.4	11.9	601	US-09-949-016-112793	Sequence 112793, A



```

C 97 47.4 11.9 601 3 US-09-949-016-112875 Sequence 112875,
C 98 47.4 11.9 601 3 US-09-949-016-112877 Sequence 112877,
C 99 47.4 11.9 601 3 US-09-949-016-112851 Sequence 112951,
C 100 47.4 11.9 601 3 US-09-949-016-112953 Sequence 112953,
C 101 47.4 11.9 601 3 US-09-949-016-113037 Sequence 113037,
C 102 47.4 11.9 601 3 US-09-949-016-113039 Sequence 113039,
C 103 47.4 11.9 106380 3 US-09-949-016-17553 Sequence 17553, A
C 104 47.4 11.9 127280 3 US-09-949-016-14857 Sequence 14857, A
C 105 47.4 11.9 122266 3 US-09-949-016-14860 Sequence 14860, A
C 106 47.4 11.9 150833 3 US-09-949-016-14859 Sequence 14859, A
C 107 47.4 11.9 159963 3 US-09-949-016-14858 Sequence 14858, A
C 108 47.4 11.9 171130 3 US-09-949-016-14861 Sequence 14861, A
C 109 47.4 11.9 256287 3 US-09-949-016-14668 Sequence 14668, A
C 110 47.4 11.9 360470 3 US-09-949-016-13173 Sequence 13173, A
C 111 47.2 11.8 60417 3 US-09-949-016-13312 Sequence 13312, A
C 112 47.2 11.8 321022 3 US-09-949-016-11852 Sequence 11852, A
C 113 47.2 11.8 321022 3 US-09-949-016-14166 Sequence 14166, A
C 114 47.2 11.8 321022 3 US-09-949-016-112708 Sequence 112708, A
C 115 47.2 11.8 601 3 US-09-949-016-112792 Sequence 112792,
C 116 47.2 11.8 601 3 US-09-949-016-112876 Sequence 112876,
C 117 47.2 11.8 601 3 US-09-949-016-112852 Sequence 112852,
C 118 47.2 11.8 601 3 US-09-949-016-113038 Sequence 113038,
C 119 47.2 11.8 23445 3 US-09-949-016-12955 Sequence 12955, A
C 120 47.2 11.8 818128 3 US-09-949-016-14546 Sequence 14546, A
C 121 47.2 11.8 818128 3 US-09-949-016-14547 Sequence 14547, A
C 122 47.2 11.8 818128 3 US-09-949-016-14548 Sequence 14548, A
C 123 47.2 11.8 818128 3 US-09-949-016-14549 Sequence 14549, A
C 124 47.2 11.8 818128 3 US-09-949-016-14550 Sequence 14550, A
C 125 47.2 11.8 818128 3 US-09-949-016-14551 Sequence 14551, A
C 126 47.2 11.8 818128 3 US-09-949-016-14552 Sequence 14552, A
C 127 47.2 11.8 818128 3 US-09-949-016-14553 Sequence 14553, A
C 128 47.2 11.8 818128 3 US-09-949-016-14554 Sequence 14554, A
C 129 47.2 11.8 818128 3 US-09-949-016-14555 Sequence 14555, A
C 130 47.2 11.8 818128 3 US-09-949-016-14556 Sequence 14556, A
C 131 47.2 11.8 818128 3 US-09-949-016-14557 Sequence 14557, A
C 132 47.2 11.8 818128 3 US-09-949-016-14558 Sequence 14558, A
C 133 47.2 11.8 818128 3 US-09-949-016-14559 Sequence 14559, A
C 134 47.2 11.8 818128 3 US-09-949-016-14560 Sequence 14560, A
C 135 47.2 11.8 818128 3 US-09-949-016-14561 Sequence 14561, A
C 136 47.2 11.8 818128 3 US-09-949-016-14562 Sequence 14562, A
C 137 47.2 11.8 818128 3 US-09-949-016-14564 Sequence 14564, A
C 138 47.2 11.8 818128 3 US-09-949-016-14565 Sequence 14565, A
C 139 47.2 11.8 818128 3 US-09-949-016-14566 Sequence 14566, A
C 140 47.2 11.8 818128 3 US-09-949-016-14567 Sequence 14567, A
C 141 46.8 11.7 601 3 US-09-949-016-24727 Sequence 24727, A
C 142 46.8 11.7 601 3 US-09-949-016-82381 Sequence 82381, A
C 143 46.8 11.7 601 3 US-09-949-016-149790 Sequence 149790,
C 144 46.8 11.7 601 3 US-09-949-016-187551 Sequence 187551,
C 145 46.8 11.7 78810 3 US-09-949-016-16198 Sequence 16198, A
C 146 46.8 11.7 146401 3 US-09-949-016-16151 Sequence 16151, A
C 147 46.8 11.7 167708 3 US-09-949-016-16423 Sequence 16423, A
C 148 46.8 11.7 194790 3 US-09-949-016-15393 Sequence 15393, A
C 149 46.6 11.7 601 3 US-09-949-016-60589 Sequence 60589, A
C 150 46.6 11.7 601 3 US-09-949-016-60590 Sequence 60590, A

```

## ALIGNMENTS

```

RESULT 1
US-09-949-016-122917/c
; Sequence 122917, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```

```

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 122917
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-122917

```

```

Query Match 13.1%; Score 52.4; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 4.5e-06;
Matches 102; Conservative 2; Mismatches 63; Indels 3; Gaps 1;

```

```

QY 88 CCTGTGAAAGTATGATGACAGGCAAGTACCTTTTGTACTGTTGATATAT 147
DB 392 CATCAGGGCAAGAGAGATTTGAAAGCTTCACTTCACTTGTGTTTCTATAT 333
QY 148 ATATATACATACACACACACACACACACACACACACACACACACACAC 207
DB 332 ATATATACACACACACACACACACACACACACACACACACACACACAC 276
QY 208 TACACCACTATTTCACCACTATCTATCTCAGAGTAAGACATAA 257
DB 275 TATTCATATATGGAATATATATATATATATATATATATATATATATGA 226

```

```

RESULT 2
US-09-949-016-123011/c
; Sequence 123011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 123011
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123011

```

```

Query Match 13.1%; Score 52.4; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 4.5e-06;
Matches 102; Conservative 2; Mismatches 63; Indels 3; Gaps 1;

```

```

QY 88 CCTGTGAAAGTATGATGACAGGCAAGTACCTTTTGTACTGTTGATATAT 147
DB 392 CATCAGGGCAAGAGAGATTTGAAAGCTTCACTTCACTTGTGTTTCTATAT 333
QY 148 ATATATACATACACACACACACACACACACACACACACACACACACAC 207
DB 332 ATATATACACACACACACACACACACACACACACACACACACACACAC 276
QY 208 TACACCACTATTTCACCACTATCTATCTCAGAGTAAGACATAA 257
DB 275 TATTCATATATGGAATATATATATATATATATATATATATATATATGA 226

```

```

RESULT 3
US-09-949-016-123105/c
; Sequence 123105, Application US/09949016

```



Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 123105  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123105

Query Match  
Best Local Similarity 13.1%; Score 52.4; DB 3; Length 601;  
Matches 102; Conservative 2; Mismatches 63; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGCGAAGTGGCTTCTTTTGTACTGTTGATATAT 147  
392 CATGAGGCAAGAGACAGATTTGAAAGCTTCTGCTGTTGGTGTATTTCTATATAT 333

148 ATATATACATACACACACACACACACACACACACACATATTTTAACTCGGGCATTAATG 207  
332 ATATATACACACACACACACACACACACACACACATATATATATATATACAC--ACACATATA 276

208 TACACCACTATTTCAAAGTAAACTACTATCTCGAAGTAAGACATPA 257  
275 TATTCATATATGGAATATGTATATATATATATAGAAATATATAGA 226

RESULT 4  
US-09-949-016-123199/c  
Sequence 123199, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 123199  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123199

Query Match  
Best Local Similarity 13.1%; Score 52.4; DB 3; Length 601;  
Matches 102; Conservative 2; Mismatches 63; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGCGAAGTGGCTTCTTTTGTACTGTTGATATAT 147  
392 CATGAGGCAAGAGACAGATTTGAAAGCTTCTGCTGTTGGTGTATTTCTATATAT 333

148 ATATATACATACACACACACACACACACACACACACATATTTTAACTCGGGCATTAATG 207  
332 ATATATACACACACACACACACACACACACACACATATATATATATATACAC--ACACATATA 276

208 TACACCACTATTTCAAAGTAAACTACTATCTCGAAGTAAGACATPA 257  
275 TATTCATATATGGAATATGTATATATATATATAGAAATATATAGA 226

332 ATATATACACACACACACACACACACACACACACATATATATATATATACAC--ACACATATA 276

208 TACACCACTATTTCAAAGTAAACTACTATCTCGAAGTAAGACATPA 257  
275 TATTCATATATGGAATATGTATATATATATATAGAAATATATAGA 226

RESULT 5  
US-09-949-016-123293/c  
Sequence 123293, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 123293  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123293

Query Match  
Best Local Similarity 13.1%; Score 52.4; DB 3; Length 601;  
Matches 102; Conservative 2; Mismatches 63; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGCGAAGTGGCTTCTTTTGTACTGTTGATATAT 147  
392 CATGAGGCAAGAGACAGATTTGAAAGCTTCTGCTGTTGGTGTATTTCTATATAT 333

148 ATATATACATACACACACACACACACACACACACACATATTTTAACTCGGGCATTAATG 207  
332 ATATATACACACACACACACACACACACACACACATATATATATATATACAC--ACACATATA 276

208 TACACCACTATTTCAAAGTAAACTACTATCTCGAAGTAAGACATPA 257  
275 TATTCATATATGGAATATGTATATATATATATAGAAATATATAGA 226

RESULT 6  
US-10-009-332-31/c  
Sequence 31, Application US/10009332  
Patent No. 6716613  
GENERAL INFORMATION:  
APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.  
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING AGGRECANASE ACTIVITY  
FILE REFERENCE: 067541  
CURRENT APPLICATION NUMBER: US/10/009,332  
PRIOR FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: JPA Hei 11-321740  
PRIOR FILING DATE: 1999-11-11  
PRIOR APPLICATION NUMBER: JPA 2000-144020  
PRIOR FILING DATE: 2000-05-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 31  
LENGTH: 3455  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: promoter















```
US-09-949-016-123202/c
: Sequence 123202, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 123202
: LENGTH: 601
: TYPE: DNA
: ORGANISM: Human
US-09-949-016-123202

Query Match      12.8%; Score 51.2; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY 88 CCTGTGAAAGTAGTATGACGAGGCAAGTGGCCCTACTTTTGTACTGTGTGATATAT 147
DB 405 CATCAGGGCAAGAGACAGATTGTAAGCTACCTCTCAGTTGGGTGTATTCTATATAT 346
QY 148 ATATATATACATACACACACACACACACACACATATATTTTAACTCGGACATTAATG 207
DB 345 ATATATATACACACACACACACACACACATATATATATATATATACAC--ACACATATA 289
QY 208 TACACCACTATTTCAAAGTAAAACTACTATCTCAGAGTAAAGACATTA 257
DB 288 TATTCATATATGAAATATGTATATATATATAGAAATATAGAAATATATGA 239

RESULT 18
US-09-949-016-123204/c
: Sequence 123204, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 123204
: LENGTH: 601
: TYPE: DNA
: ORGANISM: Human
US-09-949-016-123204

Query Match      12.8%; Score 51.2; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;
```

```
QY 148 ATATATATACATACACACACACACACACACATATATTTTAACTCGGACATTAATG 207
DB 386 ATATATATACACACACACACACACACATATATATATATATATACAC--ACACATATA 330
QY 208 TACACCACTATTTCAAAGTAAAACTACTATCTCAGAGTAAAGACATTA 257
DB 329 TATTCATATATGAAATATGTATATATATATATAGAAATATAGAAATATATGA 280

RESULT 19
US-09-949-016-123295/c
: Sequence 123295, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 123295
: LENGTH: 601
: TYPE: DNA
: ORGANISM: Human
US-09-949-016-123295

Query Match      12.8%; Score 51.2; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY 88 CCTGTGAAAGTAGTATGACGAGGCAAGTGGCCCTACTTTTGTACTGTGTGATATAT 147
DB 404 CATCAGGGCAAGAGACAGATTGTAAGCTACCTCTCAGTTGGGTGTATTCTATATAT 345
QY 148 ATATATATACATACACACACACACACACACACATATATTTTAACTCGGACATTAATG 207
DB 344 ATATATATACACACACACACACACACACATATATATATATATATACAC--ACACATATA 288
QY 208 TACACCACTATTTCAAAGTAAAACTACTATCTCAGAGTAAAGACATTA 257
DB 287 TATTCATATATGAAATATGTATATATATATATAGAAATATAGAAATATATGA 238

RESULT 20
US-09-949-016-123296/c
: Sequence 123296, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 123296
: LENGTH: 601
```



```

; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123298
Query Match      12.8%; Score 51.2; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY CCTCTGGAAGTAGTATGACAGGCGCTACCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTTTAAAGCTACCTCTCAGTTGGGTGTTATCTATATAT 346
QY ATATATACATACACACACACACACACACACACACACATATTTTAACTCGGACATAATG 207
DB ATATATACACACACACACACACACACACACACATATATATATATATACAC---ACACATATA 289
QY 208 TACACCACTTATTCAGTAAGTAAACTCTATCTCAGAGTAAGACATTA 257
DB 288 TATTCATATATGGAATATGTATATATATATAGAAATATATAGA 239

RESULT 21
US-09-949-016-123298/c
; Sequence 123298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123298
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123298
Query Match      12.8%; Score 51.2; DB 3; Length 601;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY CCTCTGGAAGTAGTATGACAGGCGCTACCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTTTAAAGCTACCTCTCAGTTGGGTGTTATCTATATAT 387
QY 148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATAATG 207
DB 386 ATATATACACACACACACACACACACACACATATATATATATATACAC---ACACATATA 330
QY 208 TACACCACTTATTCAGTAAGTAAACTCTATCTCAGAGTAAGACATTA 257
DB 329 TATTCATATATGGAATATGTATATATATATAGAAATATATAGA 280

RESULT 22
US-09-949-016-15212
; Sequence 15212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US/09/949, 016
```

```

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15212
; LENGTH: 64377
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15212
Query Match      12.8%; Score 51.2; DB 3; Length 64377;
Best Local Similarity 60.0%; Pred. No. 6.7e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY CCTCTGGAAGTAGTATGACAGGCGCTACCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTTTAAAGCTACCTCTCAGTTGGGTGTTATCTATATAT 24249
QY 148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATAATG 207
DB 24250 ATATATACACACACACACACACACACACACATATATATATATATACAC---ACACATATA 24306
QY 208 TACACCACTTATTCAGTAAGTAAACTCTATCTCAGAGTAAGACATTA 257
DB 24307 TATTCATATATGGAATATGTATATATATATAGAAATATATAGA 24356

RESULT 23
US-09-949-016-15213
; Sequence 15213, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15213
; LENGTH: 64377
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15213
Query Match      12.8%; Score 51.2; DB 3; Length 64377;
Best Local Similarity 60.0%; Pred. No. 6.7e-05;
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

QY CCTCTGGAAGTAGTATGACAGGCGCTACCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTTTAAAGCTACCTCTCAGTTGGGTGTTATCTATATAT 24249
QY 148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATAATG 207
DB 24250 ATATATACACACACACACACACACACACACATATATATATATATACAC---ACACATATA 24306
QY 208 TACACCACTTATTCAGTAAGTAAACTCTATCTCAGAGTAAGACATTA 257
DB 24307 TATTCATATATGGAATATGTATATATATATAGAAATATATAGA 24356
```



RESULT 24  
US-09-949-016-15214  
Sequence 15214, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15214  
LENGTH: 64377  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15214

Query Match  
Best Local Similarity 12.8%; Score 51.2; DB 3; Length 64377;  
Best Local Similarity 60.0%; Pred. No. 6.7e-05;  
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

Db  
24190 CATCGGGCAAGAGGACGATTGAAAGCTACCTCTGAGTTGGTGTATTCTATATAT 24249  
148 ATATATACATACACACACACACACACACACACATATATATATATATATATATAT 207  
24250 ATATATACACACACACACACACACACACATATATATATATATATATATATAT 24306  
208 TACACCACTTCTTCAAGTAAACTACTATCTCGAAGTAAAGCATTA 257  
24307 TATTCATATATGGAATATGTATATATATATATATATATATATATATATATATGA 24356

RESULT 25  
US-09-949-016-15215  
Sequence 15215, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15215  
LENGTH: 64377  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15215

Query Match  
Best Local Similarity 12.8%; Score 51.2; DB 3; Length 64377;  
Best Local Similarity 60.0%; Pred. No. 6.7e-05;  
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

Db  
24190 CATCGGGCAAGAGGACGATTGAAAGCTACCTCTGAGTTGGTGTATTCTATATAT 24249  
148 ATATATACATACACACACACACACACACACACATATATATATATATATATATAT 207  
24250 ATATATACACACACACACACACACACACATATATATATATATATATATATAT 24306  
208 TACACCACTTCTTCAAGTAAACTACTATCTCGAAGTAAAGCATTA 257  
24307 TATTCATATATGGAATATGTATATATATATATATATATATATATATATATATATGA 24356

Db  
24190 CATCGGGCAAGAGGACGATTGAAAGCTACCTCTGAGTTGGTGTATTCTATATAT 24249  
148 ATATATACATACACACACACACACACACACACATATATATATATATATATATAT 207  
24250 ATATATACACACACACACACACACACACATATATATATATATATATATATAT 24306  
208 TACACCACTTCTTCAAGTAAACTACTATCTCGAAGTAAAGCATTA 257  
24307 TATTCATATATGGAATATGTATATATATATATATATATATATATATATATATATGA 24356

RESULT 26  
US-09-949-016-15216  
Sequence 15216, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15216  
LENGTH: 64377  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15216

Query Match  
Best Local Similarity 12.8%; Score 51.2; DB 3; Length 64377;  
Best Local Similarity 60.0%; Pred. No. 6.7e-05;  
Matches 102; Conservative 1; Mismatches 64; Indels 3; Gaps 1;

Db  
24190 CATCGGGCAAGAGGACGATTGAAAGCTACCTCTGAGTTGGTGTATTCTATATAT 24249  
148 ATATATACATACACACACACACACACACACACATATATATATATATATATATAT 207  
24250 ATATATACACACACACACACACACACACATATATATATATATATATATATAT 24306  
208 TACACCACTTCTTCAAGTAAACTACTATCTCGAAGTAAAGCATTA 257  
24307 TATTCATATATGGAATATGTATATATATATATATATATATATATATATATATATGA 24356

RESULT 27  
US-10-009-332-30/C  
Sequence 30, Application US/10009332  
Patent No. 6716613  
GENERAL INFORMATION:  
APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.  
TITLE OF INVENTION: NOVEL METALLOPROTEINSE HAVING AGGRECANASE ACTIVITY  
FILE REFERENCE: Q67541  
CURRENT FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: JPA Hei 11-321740  
PRIOR FILING DATE: 1999-11-11  
PRIOR APPLICATION NUMBER: JPA 2000-144020  
PRIOR FILING DATE: 2000-05-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 30  
LENGTH: 3462



CURRENT FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: JPA Hei 11-321740

; FAULT NO. 8/18813  
; GENERAL INFORMATION:







Db 368 CATCAGGCAAGAGACAGATTGTAAGCTCTCTCAGTTGGGTGTTATTTATATAT 309  
| | | | |  
Qy 148 ATATATACATACACACACACACACACACACATATTTTAACTCGGACATTAATG 207  
| | | | |  
Db 308 ATATATATACACACACACACACACACATATATATATATATACAC---ACACATATA 252  
| | | | |  
Qy 208 TACACACCTATTCAAGTAAATACTATCTCAGAGTAAGACATTA 257  
| | | | |  
Db 251 TATTCATATATGGAATATGTATATATATATATAGAAATATAGATATATGA 202  
| | | | |

RESULT 35  
US-09-949-016-122916/c  
; Sequence 122916, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 122916  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-122916

Query Match 12.7%; Score 50.8; DB 3; Length 601;  
Best Local Similarity 59.4%; Pred. No. 1.4e-05;  
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

Qy 88 CCTCTGAAAGTAGTATGACAGGCAAGTGAAGCCCTCTTTTGTACTGTTGATATAT 147  
| | | | |  
Db 390 CATCAGGCAAGAGACAGATTGTAAGCTCTCTCAGTTGGGTGTTATTTATATAT 331  
| | | | |  
Qy 148 ATATATACATACACACACACACACACACACATATTTTAACTCGGACATTAATG 207  
| | | | |  
Db 330 ATATATATACACACACACACACACACATATATATATATATACAC---ACACATATA 274  
| | | | |  
Qy 208 TACACACCTATTCAAGTAAATACTATCTCAGAGTAAGACATTA 257  
| | | | |  
Db 273 TATTCATATATGGAATATGTATATATATATATAGAAATATAGATATATGA 224  
| | | | |

RESULT 36  
US-09-949-016-123009/c  
; Sequence 123009, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 123009

; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-123009

Query Match 12.7%; Score 50.8; DB 3; Length 601;  
Best Local Similarity 59.4%; Pred. No. 1.4e-05;  
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

Qy 88 CCTCTGAAAGTAGTATGACAGGCAAGTGAAGCCCTCTTTTGTACTGTTGATATAT 147  
| | | | |  
Db 368 CATCAGGCAAGAGACAGATTGTAAGCTCTCTCAGTTGGGTGTTATTTATATAT 309  
| | | | |  
Qy 148 ATATATACATACACACACACACACACACACATATTTTAACTCGGACATTAATG 207  
| | | | |  
Db 308 ATATATATACACACACACACACACACATATATATATATATACAC---ACACATATA 252  
| | | | |  
Qy 208 TACACACCTATTCAAGTAAATACTATCTCAGAGTAAGACATTA 257  
| | | | |  
Db 251 TATTCATATATGGAATATGTATATATATATAGAAATATAGATATATGA 202  
| | | | |

RESULT 37  
US-09-949-016-123010/c  
; Sequence 123010, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 123010  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-123010

Query Match 12.7%; Score 50.8; DB 3; Length 601;  
Best Local Similarity 59.4%; Pred. No. 1.4e-05;  
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

Qy 88 CCTCTGAAAGTAGTATGACAGGCAAGTGAAGCCCTCTTTTGTACTGTTGATATAT 147  
| | | | |  
Db 390 CATCAGGCAAGAGACAGATTGTAAGCTCTCTCAGTTGGGTGTTATTTATATAT 331  
| | | | |  
Qy 148 ATATATACATACACACACACACACACACACATATTTTAACTCGGACATTAATG 207  
| | | | |  
Db 330 ATATATATACACACACACACACACACATATATATATATATACAC---ACACATATA 274  
| | | | |  
Qy 208 TACACACCTATTCAAGTAAATACTATCTCAGAGTAAGACATTA 257  
| | | | |  
Db 273 TATTCATATATGGAATATGTATATATATATAGAAATATAGATATATGA 224  
| | | | |

RESULT 38  
US-09-949-016-123103/c  
; Sequence 123103, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307



```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123103
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123103

Query Match
Best Local Similarity 12.7%; Score 50.8; DB 3; Length 601;
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGAGGAGGCTTCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTGAAAGCTCCTCTCAGTTGGGTGTTATTTCTATATAT 309
148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATTAATG 207
DB ATATATATACACACACACACACACACACATATATATATATATATACAC--ACACATATA 252
208 TACACCACTTATTCAGTAAGTAAACTACTATCTCAGAGTAAGACATATA 257
DB TATTCATATATGGAATATGATATATATATATATAGAAATATAGAAATATATAGA 202

RESULT 39
US-09-949-016-123104/c
; Sequence 123104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123104
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123104

Query Match
Best Local Similarity 12.7%; Score 50.8; DB 3; Length 601;
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGAGGAGGCTTCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTGAAAGCTCCTCTCAGTTGGGTGTTATTTCTATATAT 331
148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATTAATG 207
DB ATATATATACACACACACACACACACACATATATATATATATATACAC--ACACATATA 274
208 TACACCACTTATTCAGTAAGTAAACTACTATCTCAGAGTAAGACATATA 257
DB TATTCATATATGGAATATGATATATATATATATAGAAATATAGAAATATATAGA 224
```

```

RESULT 40
US-09-949-016-123197/c
; Sequence 123197, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123197
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123197

Query Match
Best Local Similarity 12.7%; Score 50.8; DB 3; Length 601;
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;

88 CCTCTGGAAGTAGTATGACAGGAGGAGGCTTCTTTTGTACTGTTGATATAT 147
DB CATCAGGGCAAGAGACAGATTGAAAGCTCCTCTCAGTTGGGTGTTATTTCTATATAT 309
148 ATATATACATACACACACACACACACACACACACATATTTTAACTCGGACATTAATG 207
DB ATATATATACACACACACACACACACACATATATATATATATATACAC--ACACATATA 252
208 TACACCACTTATTCAGTAAGTAAACTACTATCTCAGAGTAAGACATATA 257
DB TATTCATATATGGAATATGATATATATATATATAGAAATATAGAAATATATAGA 202
```

```

RESULT 41
US-09-949-016-123198/c
; Sequence 123198, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123198
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-123198

Query Match
Best Local Similarity 12.7%; Score 50.8; DB 3; Length 601;
Matches 101; Conservative 2; Mismatches 64; Indels 3; Gaps 1;
```







PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 123015  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123015

Query Match 12.5%; Score 50; DB 3; Length 601;  
Best Local Similarity 65.7%; Pred. No. 2.4e-05;  
Matches 71; Conservative 1; Mismatches 36; Indels 0; Gaps 0;

QY 88 CCTCTGGAAGTATGACAGGCGCAAGTGAAGCCCTACTTTTGTACTGTTGATATAT 147  
DB 413 CATCAGGGCAAGAGACAGATTGAAAGCTACTCTCAGTTGGGTGTTATTTCTATATAT 354  
QY 148 ATATATACATACACACACACACACACACACACACATATATTTTAATC 195  
DB 353 ATATATACACACACACACACACACACACACATATATATATATATACAC 306

RESULT 46  
US-09-949-016-123109/c  
Sequence 123109, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 123109  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123109

Query Match 12.5%; Score 50; DB 3; Length 601;  
Best Local Similarity 65.7%; Pred. No. 2.4e-05;  
Matches 71; Conservative 1; Mismatches 36; Indels 0; Gaps 0;

QY 88 CCTCTGGAAGTATGACAGGCGCAAGTGAAGCCCTACTTTTGTACTGTTGATATAT 147  
DB 413 CATCAGGGCAAGAGACAGATTGAAAGCTACTCTCAGTTGGGTGTTATTTCTATATAT 354  
QY 148 ATATATACATACACACACACACACACACACACACATATATTTTAATC 195  
DB 353 ATATATACACACACACACACACACACACACATATATATATATATACAC 306

RESULT 47  
US-09-949-016-123203/c  
Sequence 123203, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 123203  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123203

FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 123203  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123203

Query Match 12.5%; Score 50; DB 3; Length 601;  
Best Local Similarity 65.7%; Pred. No. 2.4e-05;  
Matches 71; Conservative 1; Mismatches 36; Indels 0; Gaps 0;

QY 88 CCTCTGGAAGTATGACAGGCGCAAGTGAAGCCCTACTTTTGTACTGTTGATATAT 147  
DB 413 CATCAGGGCAAGAGACAGATTGAAAGCTACTCTCAGTTGGGTGTTATTTCTATATAT 354  
QY 148 ATATATACATACACACACACACACACACACACACATATATTTTAATC 195  
DB 353 ATATATACACACACACACACACACACACACATATATATATATATACAC 306

RESULT 48  
US-09-949-016-123297/c  
Sequence 123297, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: CLO01307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 123297  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-123297

Query Match 12.5%; Score 50; DB 3; Length 601;  
Best Local Similarity 65.7%; Pred. No. 2.4e-05;  
Matches 71; Conservative 1; Mismatches 36; Indels 0; Gaps 0;

QY 88 CCTCTGGAAGTATGACAGGCGCAAGTGAAGCCCTACTTTTGTACTGTTGATATAT 147  
DB 413 CATCAGGGCAAGAGACAGATTGAAAGCTACTCTCAGTTGGGTGTTATTTCTATATAT 354  
QY 148 ATATATACATACACACACACACACACACACACACATATATTTTAATC 195  
DB 353 ATATATACACACACACACACACACACACACATATATATATATATACAC 306

RESULT 49  
US-09-526-193A-23/c  
Sequence 23, Application US/09526193A  
Patent No. 6617122  
GENERAL INFORMATION:







GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2006, 16:04:09 ; Search time 1439 Seconds

(without alignments) 3415.604 Million cell updates/sec

Title: SEQ1-33670G

Perfect score: 399.2

Sequence: 1.aagatgctctctctcctccta.....taactcagatcagaacgag 400

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues 37784340

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Published Applications NA Main:

1: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US07\_PUBCOMB.seq.\*

2: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US08\_PUBCOMB.seq.\*

3: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US09\_PUBCOMB.seq.\*

4: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*

5: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*

6: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10A\_PUBCOMB.seq.\*

7: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10A\_PUBCOMB.seq.\*

8: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10C\_PUBCOMB.seq.\*

9: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10D\_PUBCOMB.seq.\*

10: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10E\_PUBCOMB.seq.\*

11: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10F\_PUBCOMB.seq.\*

12: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US10G\_PUBCOMB.seq.\*

13: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US11A\_PUBCOMB.seq.\*

14: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US11B\_PUBCOMB.seq.\*

15: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US11C\_PUBCOMB.seq.\*

16: /EMC\_Celerra\_SIDS3/prodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397.6	99.6	504	US-09-925-065A-153932	Sequence 153932,
2	397.6	99.6	504	US-09-925-065A-153933	Sequence 153933,
3	397.6	99.6	504	US-09-925-065A-153933	Sequence 153933,
4	397.6	99.6	504	US-09-925-065A-153933	Sequence 153933,
5	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
6	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
7	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
8	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
9	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
10	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
11	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
12	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
13	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
14	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
15	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
16	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,
17	397.6	99.6	509	US-10-301-480-247520	Sequence 247520,

18	54.2	13.6	599	US-09-925-065A-321660	Sequence 321660,
19	54.2	13.6	599	US-09-925-065A-321660	Sequence 321660,
20	54.2	13.6	616	US-10-301-480-396282	Sequence 396282,
21	54.2	13.6	616	US-10-301-480-396282	Sequence 396282,
22	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
23	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
24	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
25	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
26	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
27	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
28	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
29	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
30	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
31	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
32	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
33	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
34	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
35	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
36	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
37	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
38	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
39	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
40	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
41	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
42	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
43	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
44	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
45	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
46	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
47	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
48	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
49	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
50	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
51	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
52	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
53	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
54	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
55	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
56	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
57	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
58	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
59	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
60	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
61	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
62	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
63	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
64	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
65	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
66	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
67	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
68	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
69	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
70	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
71	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
72	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
73	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
74	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
75	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
76	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
77	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
78	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
79	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
80	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
81	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
82	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
83	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
84	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
85	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
86	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
87	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
88	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
89	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,
90	54.2	13.6	682	US-10-301-480-566528	Sequence 566528,



C 91	51	12.8	3464	8	US-10-763-210-26	Sequence 26, Appl
C 92	51	12.8	3467	8	US-10-763-210-25	Sequence 25, Appl
C 93	51	12.8	3467	8	US-10-763-210-29	Sequence 27, Appl
C 94	51	12.8	3468	8	US-10-763-210-27	Sequence 27, Appl
C 95	51	12.8	3470	8	US-10-763-210-28	Sequence 28, Appl
C 96	51	12.8	3473	8	US-10-763-210-24	Sequence 3, Appl1
C 97	51	12.8	28854	3	US-09-741-151-3	Sequence 642552
C 98	50.8	12.7	591	4	US-09-925-065A-642552	Sequence 642552
C 99	50.8	12.7	591	5	US-09-925-065A-642552	Sequence 642552
C 100	50.8	12.7	644	12	US-10-301-480-391371	Sequence 391371
C 101	50.8	12.7	644	12	US-10-301-480-1004780	Sequence 1004780
C 102	50.8	12.7	962	12	US-10-301-480-544114	Sequence 544114
C 103	50.8	12.7	962	12	US-10-301-480-1157523	Sequence 1157523
C 104	50.8	12.7	92726	3	US-09-997-722-193	Sequence 193, App
C 105	50.8	12.7	134945	7	US-10-085-117-155	Sequence 355, App
C 106	50.6	12.7	531	4	US-09-925-065A-486389	Sequence 486389
C 107	50.6	12.7	531	5	US-09-925-065A-486389	Sequence 486389
C 108	50.6	12.7	562	12	US-10-301-480-294578	Sequence 294578
C 109	50.6	12.7	562	12	US-10-301-480-907987	Sequence 907987
C 110	50.6	12.7	593	4	US-09-925-065A-206820	Sequence 206820
C 111	50.6	12.7	593	5	US-09-925-065A-206820	Sequence 206820
C 112	50.6	12.7	633	4	US-09-925-065A-99858	Sequence 99858, A
C 113	50.6	12.7	633	5	US-09-925-065A-99858	Sequence 99858, A
C 114	50.6	12.7	640	12	US-10-301-480-200523	Sequence 200523
C 115	50.6	12.7	640	12	US-10-301-480-813932	Sequence 813932
C 116	50.6	12.7	694	6	US-10-027-632-111223	Sequence 111223
C 117	50.6	12.7	694	7	US-10-027-632-111223	Sequence 111223
C 118	50.4	12.6	612	4	US-09-925-065A-742315	Sequence 742315
C 119	50.4	12.6	612	5	US-09-925-065A-742315	Sequence 742315
C 120	50.4	12.6	640	12	US-10-301-480-71593	Sequence 71593, A
C 121	50.4	12.6	640	12	US-10-301-480-685002	Sequence 685002
C 122	50.4	12.6	657	12	US-10-301-480-605379	Sequence 605379
C 123	50.4	12.6	657	12	US-10-301-480-1218788	Sequence 1218788
C 124	50.4	12.6	827	6	US-10-027-632-165881	Sequence 165881
C 125	50.4	12.6	827	6	US-10-027-632-165882	Sequence 165882
C 126	50.4	12.6	827	7	US-10-027-632-165881	Sequence 165881
C 127	50.4	12.6	827	7	US-10-027-632-165882	Sequence 165882
C 128	50.4	12.6	967	12	US-10-301-480-560457	Sequence 560457
C 129	50.4	12.6	967	12	US-10-301-480-1173866	Sequence 1173866
C 130	50.4	12.6	1000	12	US-10-301-480-557803	Sequence 557803
C 131	50.4	12.6	1000	12	US-10-301-480-1171212	Sequence 1171212
C 132	50.4	12.6	1994	4	US-09-925-065A-683118	Sequence 683118
C 133	50.4	12.6	1994	5	US-09-925-065A-683118	Sequence 683118
C 134	50.2	12.6	106	9	US-10-674-124A-25224	Sequence 25224, A
C 135	50.2	12.6	186	9	US-10-674-124A-24700	Sequence 24700, A
C 136	50.2	12.6	341	9	US-10-674-124A-11186	Sequence 11186, A
C 137	50.2	12.6	341	9	US-09-925-065A-156885	Sequence 156885
C 138	50.2	12.6	379	5	US-09-925-065A-156885	Sequence 156885
C 139	50.2	12.6	391	12	US-10-301-480-250077	Sequence 250077
C 140	50.2	12.6	391	12	US-10-301-480-863486	Sequence 863486
C 141	50.2	12.6	401	12	US-10-301-480-250076	Sequence 250076
C 142	50.2	12.6	401	12	US-10-301-480-863485	Sequence 863485
C 143	50.2	12.6	413	4	US-09-925-065A-156884	Sequence 156884
C 144	50.2	12.6	413	4	US-09-925-065A-156884	Sequence 156884
C 145	50.2	12.6	622	7	US-10-027-632-115412	Sequence 115412
C 146	50.2	12.6	622	7	US-10-027-632-115412	Sequence 115412
C 147	50.2	12.6	668	6	US-10-027-632-228805	Sequence 228805
C 148	50.2	12.6	668	6	US-10-027-632-228805	Sequence 228805
C 149	50.2	12.6	668	7	US-10-027-632-228805	Sequence 228805
C 150	50.2	12.6	668	7	US-10-027-632-228806	Sequence 228806

## ALIGNMENTS

RESULT 1  
 US-09-925-065A-153932/c  
 ; Sequence 153932, Application US/09925065A  
 ; Publication No. US20040181048A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; Nucleotide Polymorphisms in the Human Genome

FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US/09/925, 065A  
 ; CURRENT FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243, 096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252, 147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250, 092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261, 766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289, 846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 153932  
 ; LENGTH: 504  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-925-065A-153932

Query Match 99.6%; Score 397.6; DB 4; Length 504;  
 Best Local Similarity 99.5%; Pred. No. 2.0e-102;  
 Matches 398; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 AAGATGCTCTTCTCTATATATTTAGAGTCTTGATTAAGCTTAATAATCTGTGGCG 60  
 420 AAGATGCTCTTCTCTATATATTTAGAGTCTTGATTAAGCTTAATAATCTGTGGCG 361  
 61 TACACATGAGACGTGTGATCTGTTACCTCTGGAAGTGTGACGAGGAGTGAAG 120  
 360 TACACATGAGACGTGTGATCTGTTACCTCTGGAAGTGTGACGAGGAGTGAAG 301  
 121 CTTACCTTTTGTACCTTTGATATATATATATATACACACACACACACACACA 180  
 300 CTTACCTTTTGTACCTTTGATATATATATATATACACACACACACACACACA 241  
 181 CAAATATTTTAACTCGGACATTAATGTACACACACTTTCAAAGTAAATCTACTATCT 240  
 240 CAAATATTTTAACTCGGACATTAATGTACACACACTTTCAAAGTAAATCTACTATCT 181  
 241 CAGAAATGAAGACATTAATCTGCTATTGTCTATTAACGACGTTAAGCTGCTAACTTT 300  
 180 CAGAAATGAAGACATTAATCTGCTATTGTCTATTAACGACGTTAAGCTGCTAACTTT 121  
 301 ATGAGATTCAGTTTCCCTACTATGAAACCTGAAGGTGAATCTGTGGTTTACCTTA 360  
 120 ATGAGATTCAGTTTCCCTACTATGAAACCTGAAGGTGAATCTGTGGTTTACCTTA 61

361 CTTTGAAGACGTCAATACCTTAACCTCAAGTACGAGAG 400  
 60 CTTTGAAGACGTCAATACCTTAACCTCAAGTACGAGAG 21

RESULT 2  
 US-09-925-065A-153932/c  
 ; Sequence 153932, Application US/09925065A  
 ; Publication No. US20040181048A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; Nucleotide Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US/09/925, 065A  
 ; CURRENT FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243, 096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252, 147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250, 092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261, 766  
 ; PRIOR FILING DATE: 2001-01-16



PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 153933  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-153933

Query Match 99.6%; Score 397.6; DB 4; Length 504;  
Best Local Similarity 99.2%; Pred. No. 2.8e-102;  
Matches 397; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 60  
DB 420 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 361  
QY 61 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 120  
DB 360 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 301  
QY 121 CCTACCTTTTGTACTGTGATATATATATATATATATATATATATATATATATATAT 180  
DB 300 CCTACCTTTTGTACTGTGATATATATATATATATATATATATATATATATATATAT 241  
QY 181 CAAATATTTTAACTGGGACATAAATGTACACACACTTCAAAAGTAAATCTATCT 240  
DB 240 CAAATATTTTAACTGGGACATAAATGTACACACACTTCAAAAGTAAATCTATCT 181  
QY 241 CAGAGTAAAGACATAATCCGTATTTGTCAATTAAGTAAAGTAAAGTAAAGTAAAG 300  
DB 180 CAGAGTAAAGACATAATCCGTATTTGTCAATTAAGTAAAGTAAAGTAAAGTAAAG 121  
QY 301 ATGAGTTCAGTTTCCCTACTATGAAAACTGAAAGTTGAAATCTGGTGTAACTTAA 360  
DB 120 ATGAGTTCAGTTTCCCTACTATGAAAACTGAAAGTTGAAATCTGGTGTAACTTAA 61  
QY 361 CCTTGAAGACGTCAATACCTTAAGTCAAGATCAGAACGAG 400  
DB 60 CCTTGAAGACGTCAATACCTTAAGTCAAGATCAGAACGAG 21

## RESULT 3

US-09-925-065A-153932/c  
Sequence 153932, Application US/0925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925.065A  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 153932  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-153932

Query Match 99.6%; Score 397.6; DB 5; Length 504;

Best Local Similarity 99.5%; Pred. No. 2.8e-102;  
Matches 398; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 60  
DB 420 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 361  
QY 61 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 120  
DB 360 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 301  
QY 121 CCTACCTTTTGTACTGTGATATATATATATATATATATATATATATATATATATAT 180  
DB 300 CCTACCTTTTGTACTGTGATATATATATATATATATATATATATATATATATAT 241  
QY 181 CAAATATTTTAACTGGGACATAAATGTACACACACTTCAAAAGTAAATCTATCT 240  
DB 240 CAAATATTTTAACTGGGACATAAATGTACACACACTTCAAAAGTAAATCTATCT 181  
QY 241 CAGAGTAAAGACATAATCCGTATTTGTCAATTAAGTAAAGTAAAGTAAAGTAAAG 300  
DB 180 CAGAGTAAAGACATAATCCGTATTTGTCAATTAAGTAAAGTAAAGTAAAGTAAAG 121  
QY 301 ATGAGTTCAGTTTCCCTACTATGAAAACTGAAAGTTGAAATCTGGTGTAACTTAA 360  
DB 120 ATGAGTTCAGTTTCCCTACTATGAAAACTGAAAGTTGAAATCTGGTGTAACTTAA 61  
QY 361 CCTTGAAGACGTCAATACCTTAAGTCAAGATCAGAACGAG 400  
DB 60 CCTTGAAGACGTCAATACCTTAAGTCAAGATCAGAACGAG 21

## RESULT 4

US-09-925-065A-153933/c  
Sequence 153933, Application US/0925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925.065A  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 153933  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-153933

Query Match 99.6%; Score 397.6; DB 5; Length 504;  
Best Local Similarity 99.2%; Pred. No. 2.8e-102;  
Matches 397; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 60  
DB 420 AAGATGCTCTTCTCTATATATTACGAGTTCTTGAATAGCTTAAATACCTGTGCG 361  
QY 61 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 120  
DB 360 TACACATGAGACTGTGTATCTGGTACCTCTGGAAGTAGTATGCGGGCAAGTGAG 301







Db 33661 TACA CATGAACCTGCTGATACCTGCGTTACCTCTGGAAGTAGCTATGACGGGCAATGACG 33720















TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-321660

Query Match 13.6%; Score 54.2; DB 5; Length 599;  
Best Local Similarity 66.7%; Pred. No. 0.00011;  
Matches 78; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 76 TGAATCTGTTACCTCTGGAAAGTAGTATGACGGGCAAGTGGCCCTTCTTTTGTAC 135  
DB 142 TGAATAATCTTATTTTCAAAAATGTGATGAGATGATGCAAGATTCCATTTAAAAA 201  
QY 136 TGTGTATTA 192  
DB 202 TCAATTA 258

#### RESULT 20

US-10-301-480-396282  
Sequence 396282, Application US/10301480  
Publication No. US20060057564A1

GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 10/215,598  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: US 60/311,695  
PRIOR FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 396282  
LENGTH: 616  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-396282

Query Match 13.6%; Score 54.2; DB 12; Length 616;  
Best Local Similarity 66.7%; Pred. No. 0.00011;  
Matches 78; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 76 TGAATCTGTTACCTCTGGAAAGTAGTATGACGGGCAAGTGGCCCTTCTTTTGTAC 135  
DB 159 TGAATAATCTTATTTTCAAAAATGTGATGAGATGATGCAAGATTCCATTTAAAAA 218  
QY 136 TGTGTATTA 192  
DB 219 TCAATTA 275

#### RESULT 21

US-10-301-480-1009691  
Sequence 1009691, Application US/10301480  
Publication No. US20060057564A1

GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 10/215,598  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: US 60/311,695  
PRIOR FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1009691  
LENGTH: 616  
TYPE: DNA

ORGANISM: Homo sapien  
US-10-301-480-1009691

Query Match 13.6%; Score 54.2; DB 12; Length 616;  
Best Local Similarity 66.7%; Pred. No. 0.00011;  
Matches 78; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 76 TGAATCTGTTACCTCTGGAAAGTAGTATGACGGGCAAGTGGCCCTTCTTTTGTAC 135  
DB 159 TGAATAATCTTATTTTCAAAAATGTGATGAGATGATGCAAGATTCCATTTAAAAA 218  
QY 136 TGTGTATTA 192  
DB 219 TCAATTA 275

#### RESULT 22

US-10-301-480-566528/c  
Sequence 566528, Application US/10301480  
Publication No. US20060057564A1

GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 10/215,598  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: US 60/311,695  
PRIOR FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 566528  
LENGTH: 682  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-566528

Query Match 13.5%; Score 54; DB 12; Length 682;  
Best Local Similarity 59.2%; Pred. No. 0.00013;  
Matches 90; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

QY 113 AAGTAGGCTTACCTCTTTTGTACTGTTGATATATATATATATATATATATATTA 172  
DB 525 ACGGTAGGACATTA 466  
QY 173 CACACACACATTA 232  
DB 465 CACACACACACACACACACACATTTCTTAAAGATTATACACAAATCATATGAAATCC 406  
QY 233 TACTATCTCAGAGTAAGATTAAGATATATATATATATATATATATATATATTA 264  
DB 405 TTCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 374

#### RESULT 23

US-10-301-480-1179937/c  
Sequence 1179937, Application US/10301480  
Publication No. US20060057564A1

GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 10/215,598  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: US 60/311,695  
PRIOR FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0







Db 150 ATATTATATATACACAGTATTTTATATATATATATATATATATATATATATATA 91







GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
PRIORITY FILING DATE: 2002-11-21  
PRIORITY APPLICATION NUMBER: US 10/215,598  
PRIORITY FILING DATE: 2002-08-09  
PRIORITY APPLICATION NUMBER: US 60/311,695  
PRIORITY FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 132528  
LENGTH: 578  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-132528

Query Match 13.3%; Score 53.2; DB 12; Length 578;  
Best Local Similarity 58.7%; Pred. No. 0.00021;  
Matches 88; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

95 AAAGTAGGTATGCGAGGCGAGGCGCTTACTCTTTGTTGATGTTGATATATATATA 154  
150 ATATTATATATACACAGTATTTTATATATATATATATATATATATATATATA 91  
155 CATACCA 214  
90 TAT 31  
215 CCTATTCAAGTAAAGTAACTACTATCTCAGA 244  
30 CAAAACAAAGTAAAGTTACTCTCTCGA 1

## RESULT 34

US-10-301-480-745936/c  
Sequence 745936, Application US/10301480  
Publication No. US20060057564A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
PRIORITY FILING DATE: 2002-11-21  
PRIORITY APPLICATION NUMBER: US 10/215,598  
PRIORITY FILING DATE: 2002-08-09  
PRIORITY APPLICATION NUMBER: US 60/311,695  
PRIORITY FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 745936  
LENGTH: 578  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-745936

Query Match 13.3%; Score 53.2; DB 12; Length 578;  
Best Local Similarity 58.7%; Pred. No. 0.00021;  
Matches 88; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

95 AAAGTAGGTATGCGAGGCGAGGCGCTTACTCTTTGTTGATGTTGATATATATATA 154  
150 ATATTATATATACACAGTATTTTATATATATATATATATATATATATATATA 91  
155 CATACCA 214  
90 TAT 31  
215 CCTATTCAAGTAAAGTAACTACTATCTCAGA 244  
111 GCAAGTAGGCGCTTCTTTGTTGATGTTGATATATATATATATATATATATATAT 170  
236 GAAAGAGGCCCATCTCTGTTGCTGATATATATATATATATATATATATATATAT 177

Db 30 CAAAACAAAGTAAAGTTACTCTCTCGA 1

RESULT 35  
US-10-301-480-745937/c  
Sequence 745937, Application US/10301480  
Publication No. US20060057564A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
PRIORITY FILING DATE: 2002-11-21  
PRIORITY APPLICATION NUMBER: US 10/215,598  
PRIORITY FILING DATE: 2002-08-09  
PRIORITY APPLICATION NUMBER: US 60/311,695  
PRIORITY FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 745937  
LENGTH: 578  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-745937

Query Match 13.3%; Score 53.2; DB 12; Length 578;  
Best Local Similarity 58.7%; Pred. No. 0.00021;  
Matches 88; Conservative 2; Mismatches 60; Indels 0; Gaps 0;

95 AAAGTAGGTATGCGAGGCGAGGCGCTTACTCTTTGTTGATGTTGATATATATATA 154  
150 ATATTATATATACACAGTATTTTATATATATATATATATATATATATATATA 91  
155 CATACCA 214  
90 TAT 31  
215 CCTATTCAAGTAAAGTAACTACTATCTCAGA 244  
30 CAAAACAAAGTAAAGTTACTCTCTCGA 1

RESULT 36  
US-10-301-480-411688/c  
Sequence 411688, Application US/10301480  
Publication No. US20060057564A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
FILE REFERENCE: 108827.137  
CURRENT APPLICATION NUMBER: US/10/301,480  
PRIORITY FILING DATE: 2002-11-21  
PRIORITY APPLICATION NUMBER: US 10/215,598  
PRIORITY FILING DATE: 2002-08-09  
PRIORITY APPLICATION NUMBER: US 60/311,695  
PRIORITY FILING DATE: 2001-08-10  
NUMBER OF SEQ ID NOS: 1226818  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 411688  
LENGTH: 535  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-301-480-411688

Query Match 13.3%; Score 53; DB 12; Length 535;  
Best Local Similarity 61.5%; Pred. No. 0.00023;  
Matches 83; Conservative 1; Mismatches 51; Indels 0; Gaps 0;

111 GCAAGTAGGCGCTTCTTTGTTGATGTTGATATATATATATATATATATATATAT 170  
236 GAAAGAGGCCCATCTCTGTTGCTGATATATATATATATATATATATATATATAT 177



Oy	17	CACACACACACATTTTAACTCGGAGCAATAATTACACACCTATTGAAGTAAA	230
Dd	176	CACACACACACACTATGTATGGTGCTATATGATATACACACAACACTTAG	117
Oy	231	ACTACTATCTCAGAA	245
Dd	116	GCTTGCCTTTTAAAA	102

RESULT 37  
US-10-301

US-10-301-480-411689/c  
; Sequence 411689, Application US/10301480  
; Publication No. US20060057564A1  
; PUBLICATION INFORMATION

APPLICANT: Wang, David G.  
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms  
 TITLE OF INVENTION: In the Human Genome  
 FILE REFERENCE: 108827.137  
 CURRENT APPLICATION NUMBER: US/10/301,480  
 CURRENT FILING DATE: 2002-11-21  
 PRIOR APPLICATION NUMBER: US 10/215,598  
 PRIOR FILING DATE: 2002-08-09  
 PRIOR APPLICATION NUMBER: US 60/311,695  
 PRIOR FILING DATE: 2001-08-10  
 NUMBER OF SEQ ID NOS: 1226818  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 411689  
 LENGTH: 535  
 TYPE: DNA  
 ORGANISM: Homo sapien  
 US-10-301-480-411689

Query Match	13.3%	Score 53	DB 12	Length 535
Best Local Similarity	61.5%	Pred. NO. 0.00023		
Matches 83	Conservative 1	Mismatches 51	Indels 0	Gaps 0

Oy	111	GCAGATGAGGCCGCTACCTTTTGGACGTTTGATATATATATATATACATACACACACACA	170
Db	236	GGAAGAGGCCCACTCTCTGCTTGCGATATATATATATATATATACACACACACACACA	177
Oy	171	CACACACACACAYATATTTTAACTGGGGAATTAATGTGACCCACTATTCAAAGTAAAA	230
Db	176	CACACACACACACACATATGATATGTTGTGTGATATATATATACACAAACAACTTATAG	117
Oy	231	ACTACTATCTCAGAA	245
Db	116	GCTTGCTTTTAAAA	102

RESULT 38  
US-10-301

US-10-301-480-1025097/c  
; Sequence 1025097, Application US/10301480  
; Publication No. US20060057564A1

```

1  APPLICANT: Wang, David G.
2  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
3  TITLE OF INVENTION: in the Human Genome
4  FILE REFERENCE: 108827.137
5  CURRENT APPLICATION NUMBER: US/10/301,480
6  CURRENT FILING DATE: 2002-11-21
7  PRIOR APPLICATION NUMBER: US 10/215,598
8  PRIOR FILING DATE: 2002-08-09
9  PRIOR APPLICATION NUMBER: US 60/311,695
10 PRIOR FILING DATE: 2001-08-10
11 NUMBER OF SEQ ID NOS: 1226818
12 SOFTWARE: FastSeq for Windows Version 4.0
13 SEQ ID NO 1025097

```

Query Match	13.3%	Score 53;	DB 12;	Length 535;
Best Local Similarity	61.5%;	Pred. No. 0.00023;		
Matches	83;	Conservative	1;	Mismatches 51;
				Indels 0;
				Gaps 0;

[illegible]

RESULT 39

US-10-301-480-1025098/c  
; Sequence 1025098, Application US/10301480  
; Publication No. US20060057564A1

```

# APPLICANT: Wang, David G.
# TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
# TITLE OF INVENTION: in the Human Genome
# FILE REFERENCE: 108827.137
# CURRENT APPLICATION NUMBER: US/10/301,480
# CURRENT FILING DATE: 2002-11-21
# PRIOR APPLICATION NUMBER: US 10/215,598
# PRIOR FILING DATE: 2002-08-09
# PRIOR APPLICATION NUMBER: US 60/311,695
# PRIOR FILING DATE: 2001-08-10
# NUMBER OF SEQ ID NOS: 126818
# SOFTWARE: FastSeq for Windows Version 4.0
# SEQ ID NO 1025098
# LENGTH: 535
# TYPE: DNA
# ORGANISM: Homo sapien
# US-10-301-480-1025098

```

Query Match	13.3%	Score 53	DB 12	Length 535
Best Local Similarity	61.5%	Pred. No. 0.00023		
Matches	83	Conservative	1	Mismatches 51; Indels 0; Gaps 0

Qy	111	GGAAGGAGGCCCTACTCTTTTGTACTGTTGATATATATATATACATACACACACACA	170
Db	236	GGAAGAGGCCCATCTCTTGTCTGATATATATATATATATATATACACACACACACA	177
Qy	171	CACACACACACATATTTTAACTCGGGACATTAATGTACACCACTATTCAAGTAA	230
Db	176	CACACACACACACATATATATGTTGTGTATATATATATATACACACACATCTATAG	117
Qy	231	ACTACTATCTCAGAA	245
Db	116	GCTTGCTTTTAAAA	102

RESULT 4C  
US-10-995

US-10-995-561-26459/c  
; Sequence 26459, Application US/10995561  
; Publication No. US20050272054A1

```

? APPLICANT: CARGILL, Michele et al.
? TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
? TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
? TITLE OF INVENTION: DETECTION AND USES THEREOF
? FILE REFERENCE: CL001559
? CURRENT APPLICATION NUMBER: US/10/995,561
? CURRENT FILING DATE: 2004-11-24
? NUMBER OF SEQ ID NOS: 85702
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 25459
? LENGTH: 201

```







NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 664474  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-664474

Query Match 13.2%; Score 52.6; DB 5; Length 504;  
Best Local Similarity 59.4%; Pred. No. 0.00029;  
Matches 85; Conservative 2; Mismatches 56; Indels 0; Gaps 0;

QY 53 CTGTGGCGTACATGAGTGTGATCTGTTACCTCTGGAAGTAGTATGACAGGC 112  
DB 439 CTGTGCATCTCAGATGTCCTGATTTTCTGAATTAACCTTCGTCACCTGGTTGTTGGCC 380  
QY 113 AAGTAGGCGCTACCTTTTGTACTGTTGATATATATATATATACACACACACACA 172  
DB 379 CATTGCAAGCCACTGCTACTGATTTGCAAAACACACACACACACACACACACACA 320  
QY 173 CACACACACAYATTTTAACTC 195  
DB 319 CACACACACACACACTCTCTCTC 297

RESULT 44  
US-09-925-065A-664473/c

Sequence 664473, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925, 065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243, 096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252, 147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250, 092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261, 766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289, 846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 664473  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-664473

Query Match 13.2%; Score 52.6; DB 5; Length 504;  
Best Local Similarity 59.4%; Pred. No. 0.00029;  
Matches 85; Conservative 2; Mismatches 56; Indels 0; Gaps 0;

QY 53 CTGTGGCGTACATGAGTGTGATCTGTTACCTCTGGAAGTAGTATGACAGGC 112  
DB 439 CTGTGCATCTCAGATGTCCTGATTTTCTGAATTAACCTTCGTCACCTGGTTGTTGGCC 380  
QY 113 AAGTAGGCGCTACCTTTTGTACTGTTGATATATATATATATACACACACACACA 172  
DB 379 CATTGCAAGCCACTGCTACTGATTTGCAAAACACACACACACACACACACACACA 320  
QY 173 CACACACACAYATTTTAACTC 195  
DB 319 CACACACACACACACTCTCTCTC 297

RESULT 45  
US-09-925-065A-664474/c

Sequence 664474, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925, 065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243, 096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252, 147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250, 092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261, 766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289, 846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 664474  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-664474

Query Match 13.2%; Score 52.6; DB 5; Length 504;  
Best Local Similarity 59.4%; Pred. No. 0.00029;  
Matches 85; Conservative 2; Mismatches 56; Indels 0; Gaps 0;

QY 53 CTGTGGCGTACATGAGTGTGATCTGTTACCTCTGGAAGTAGTATGACAGGC 112  
DB 439 CTGTGCATCTCAGATGTCCTGATTTTCTGAATTAACCTTCGTCACCTGGTTGTTGGCC 380  
QY 113 AAGTAGGCGCTACCTTTTGTACTGTTGATATATATATATATACACACACACACA 172  
DB 379 CATTGCAAGCCACTGCTACTGATTTGCAAAACACACACACACACACACACACACA 320  
QY 173 CACACACACAYATTTTAACTC 195  
DB 319 CACACACACACACACTCTCTCTC 297

RESULT 46  
US-09-925-065A-646420/c

Sequence 646420, Application US/09925065A  
Publication No. US20040181048A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925, 065A  
CURRENT FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243, 096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252, 147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250, 092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261, 766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289, 846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 646420  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-646420











GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2006, 18:33:23 ; Search time 259 Seconds

(without alignments)  
2153.288 Million cell updates/sec

Title: SRQ1-33670G

Perfect score: 399.2  
Sequence: 1.aagatgctctctctcctccta.....taactcagatcagacgag 400

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 886355 seqs, 697127050 residues

Total number of hits satisfying chosen parameters: 1772710

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%

Listing first 150 summaries

Database : Published Applications NA New:

1: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US09\_NEW\_PUB\_seq.\*  
2: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US06\_NEW\_PUB\_seq.\*  
3: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US07\_NEW\_PUB\_seq.\*  
4: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US08\_NEW\_PUB\_seq.\*  
5: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US09\_NEW\_PUB\_seq.\*  
6: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US10\_NEW\_PUB\_seq.\*  
7: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US11\_NEW\_PUB\_seq.\*  
8: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US12\_NEW\_PUB\_seq.\*  
9: /EMC\_Celextra\_SIDS3/ptocdata/1/pubpna/US13\_NEW\_PUB\_seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	13.0	1000	US-11-266-748A-200928	Sequence 200928, A
2	50.2	12.6	1000	US-11-266-748A-197937	Sequence 197937, A
3	49.4	12.4	154394	US-11-266-748A-58517	Sequence 58517, A
4	49.4	12.3	394191	US-10-506-549-3	Sequence 3, Appl1
5	49.2	12.3	755217	US-11-266-748A-29045	Sequence 29045, A
6	48.6	12.2	163354	US-11-266-748A-23437	Sequence 23437, A
7	48.4	12.1	1000	US-11-266-748A-223388	Sequence 223388, A
8	48.4	12.1	1000	US-11-266-748A-290883	Sequence 290883, A
9	48.4	12.1	1000	US-11-266-748A-342312	Sequence 342312, A
10	48.4	12.1	1000	US-11-266-748A-402247	Sequence 402247, A
11	48.4	12.1	1000	US-11-266-748A-473963	Sequence 473963, A
12	48.4	12.1	1000	US-11-266-748A-247056	Sequence 247056, A
13	47.4	11.9	1000	US-11-266-748A-203031	Sequence 203031, A
14	47.2	11.8	1000	US-11-266-748A-282791	Sequence 282791, A
15	47.2	11.8	1000	US-11-266-748A-309431	Sequence 309431, A
16	47.2	11.8	1000	US-11-266-748A-422643	Sequence 422643, A
17	46.8	11.7	1000	US-11-266-748A-205479	Sequence 205479, A
18	46.6	11.7	376	US-11-266-748A-391612	Sequence 391612, A
19	46.6	11.7	376	US-11-266-748A-482310	Sequence 482310, A
20	46.6	11.7	1000	US-11-266-748A-281212	Sequence 281212, A
21	46.6	11.7	1000	US-11-266-748A-307852	Sequence 307852, A
22	46.6	11.7	1000	US-11-266-748A-390020	Sequence 390020, A
23	46.6	11.7	1000	US-11-266-748A-480738	Sequence 480738, A
24	46.6	11.7	3564	US-11-266-748A-25970	Sequence 25970, A

25	46.2	11.6	3537	US-11-266-748A-29487	Sequence 29487, A
26	46.2	11.6	211466	US-11-266-748A-22658	Sequence 22658, A
27	46	11.5	12423	US-10-517-441-251	Sequence 251, App
28	46	11.5	12423	US-10-517-441-525	Sequence 525, App
29	46	11.5	118384	US-11-266-748A-23687	Sequence 23687, A
30	46	11.5	634888	US-10-533-365-1	Sequence 1, Appl1
31	45.8	11.5	1000	US-11-266-748A-293635	Sequence 293635, A
32	45.8	11.5	1000	US-11-266-748A-345064	Sequence 345064, A
33	45.8	11.5	1000	US-11-266-748A-405714	Sequence 405714, A
34	45.8	11.5	1000	US-11-266-748A-476760	Sequence 476760, A
35	45.8	11.5	5823	US-11-266-748A-31555	Sequence 31555, A
36	45.8	11.5	5823	US-11-266-748A-31555	Sequence 31555, A
37	45.8	11.5	6453	US-11-266-748A-31554	Sequence 31554, A
38	45.6	11.4	1000	US-11-266-748A-117839	Sequence 117839, A
39	45.6	11.4	1000	US-11-266-748A-160003	Sequence 160003, A
40	45.6	11.4	1000	US-11-266-748A-287071	Sequence 287071, A
41	45.6	11.4	1000	US-11-266-748A-290324	Sequence 290324, A
42	45.6	11.4	1000	US-11-266-748A-338500	Sequence 338500, A
43	45.6	11.4	1000	US-11-266-748A-341753	Sequence 341753, A
44	45.6	11.4	1000	US-11-266-748A-397664	Sequence 397664, A
45	45.6	11.4	1000	US-11-266-748A-401556	Sequence 401556, A
46	45.6	11.4	1000	US-11-266-748A-468710	Sequence 468710, A
47	45.6	11.4	1000	US-11-266-748A-472602	Sequence 472602, A
48	45.6	11.4	201144	US-11-266-748A-23494	Sequence 23494, A
49	45.4	11.4	740	US-11-266-748A-53663	Sequence 53663, A
50	45.4	11.4	2037	US-11-266-748A-31003	Sequence 31003, A
51	45.2	11.3	1421559	US-11-266-748A-22808	Sequence 22808, A
52	45	11.3	38945	US-11-266-748A-24969	Sequence 24969, A
53	44.8	11.2	502	US-11-266-748A-59350	Sequence 59350, A
54	44.8	11.2	1000	US-11-266-748A-198276	Sequence 198276, A
55	44.8	11.2	42790	US-11-266-748A-23990	Sequence 23990, A
56	44.6	11.2	178403	US-11-266-748A-24242	Sequence 24242, A
57	44.6	11.2	14983	US-11-266-748A-22732	Sequence 22732, A
58	44.6	11.2	84248	US-11-347-766-77	Sequence 77, Appl1
59	44.4	11.1	1000	US-11-266-748A-197513	Sequence 197513, A
60	44.4	11.1	347503	US-11-266-748A-23474	Sequence 23474, A
61	44.2	11.1	1000	US-11-266-748A-195690	Sequence 195690, A
62	44.2	11.1	1000	US-11-266-748A-203426	Sequence 203426, A
63	44.2	11.1	96217	US-11-266-748A-59101	Sequence 59101, A
64	44.2	11.1	139947	US-11-266-748A-24987	Sequence 24987, A
65	44.2	11.1	170869	US-11-266-748A-199669	Sequence 199669, A
66	44	11.0	1000	US-11-266-748A-25150	Sequence 25150, A
67	44	11.0	1000	US-11-266-748A-281845	Sequence 281845, A
68	44	11.0	1000	US-11-266-748A-308485	Sequence 308485, A
69	44	11.0	3106	US-11-266-748A-27525	Sequence 27525, A
70	44	11.0	4639	US-11-266-748A-27825	Sequence 27825, A
71	44	11.0	26757	US-11-266-748A-24847	Sequence 24847, A
72	44	11.0	176928	US-11-266-748A-60803	Sequence 60803, A
73	43.8	11.0	809	US-11-266-748A-43552	Sequence 43552, A
74	43.8	11.0	1000	US-11-266-748A-287824	Sequence 287824, A
75	43.8	11.0	1000	US-11-266-748A-339253	Sequence 339253, A
76	43.8	11.0	1000	US-11-266-748A-396596	Sequence 396596, A
77	43.8	11.0	1000	US-11-266-748A-469642	Sequence 469642, A
78	43.8	11.0	4560	US-11-266-748A-483942	Sequence 483942, A
79	43.8	11.0	5541	US-11-266-748A-30273	Sequence 30273, A
80	43.8	11.0	5541	US-11-266-748A-56385	Sequence 56385, A
81	43.8	11.0	157868	US-11-191-644-1	Sequence 1, Appl1
82	43.6	10.9	141	US-11-222-810-16	Sequence 16, Appl1
83	43.6	10.9	1000	US-11-266-748A-114970	Sequence 114970, A
84	43.6	10.9	1000	US-11-266-748A-120013	Sequence 120013, A
85	43.6	10.9	1000	US-11-266-748A-200361	Sequence 200361, A
86	43.6	10.9	1000	US-11-266-748A-282919	Sequence 282919, A
87	43.6	10.9	1000	US-11-266-748A-309559	Sequence 309559, A
88	43.6	10.9	1000	US-11-266-748A-392519	Sequence 392519, A
89	43.6	10.9	1000	US-11-266-748A-482337	Sequence 482337, A
90	43.6	10.9	3881	US-11-266-748A-26100	Sequence 26100, A
91	43.6	10.9	135090	US-10-505-928-607	Sequence 607, App
92	43.6	10.9	144649	US-11-266-748A-59314	Sequence 59314, A
93	43.4	10.9	1000	US-11-266-748A-281299	Sequence 281299, A
94	43.4	10.9	1000	US-11-266-748A-307939	Sequence 307939, A
95	43.4	10.9	1000	US-11-266-748A-390154	Sequence 390154, A
96	43.4	10.9	1000	US-11-266-748A-480872	Sequence 480872, A
97	43.4	10.9	3407	US-11-266-748A-24366	Sequence 24366, A



```

PRIORITY APPLICATION NUMBER: EP 04105482.6
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105483.4
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105507.0
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105485.9
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105484.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: US 60/662,276
PRIOR FILING DATE: 2005-03-14
PRIOR APPLICATION NUMBER: US 60/700,293
NUMBER OF SEQ ID NOS: 48396
SOFTWARE: PatentIn version 3.3
SEQ ID NO 200928
LENGTH: 1000
TYPE: DNA
ORGANISM: Homo Sapiens
US-11-266-748A-200928

Query Match      13.0%; Score 52; DB 8; Length 1000;
Best Local Similarity    55.0%; Pred. No. 5.6e-05;
Matches 122; Conservative 1; Mismatches 96; Indels 3; Gaps 1

Cy 121 CCTACCTTTTGTACTGTTTGATATATATATATATACATACACACACACACACACACA 180
Db 787 CTTTCCTTTTAATATATATATACATATACATACACACACACACACACACACACACA 728

Cy 181 CAVATTATTTAACTCGGCAGCATTAATATGACACCACCTATTCCAAAGTAAAATCTACTATCT 240
Db 727 CACATATATATAATATATACACATATATATGATCAAC---TTCTATATAAAAATTTATATGTT 671

Cy 241 CAGAAGTAAAGACATTAATCTGCTATTTGTCAATTAATCAGAGTTAAGCTGCTAACTTT 300
Db 670 CACATGTAAACAATTTAAACAGATGTTTTTTTAAAGCAAATTAATAATATTTTAAATTTT 611

Cy 301 ATGCAGTTCAAGTTTCCCTACTATATGAAAAACCTGAAGGTTGAA 342
Db 610 CTCAGTTTCAGTTTCCCAATATGATTAATTAATTTGATGATATATA 569

RESULT 2
US-11-266-748A-197937
Sequence 197937, Application US/11266748A
Publication No. US20060134663A1
GENERAL INFORMATION:
APPLICANT: Harkin, Paul
APPLICANT: Johnston, Patrick
APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcription Microarray Technology and Methods of Using the Same
```







TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-23045

Query Match 12.3%; Score 49.2; DB 8; Length 755217;  
Best Local Similarity 57.2%; Pred. No. 0.0044;  
Matches 87; Conservative 1; Mismatches 64; Indels 0; Gaps 0;

98 GTAGGTGACGAGGCAAGTGGAGGCTCTTTGTCCTGTTGATATATATATACAT 157  
DB GCATTTTATATATGCACTTCATGCAATATGATGTCGTAATATATATATATATC 87212  
158 AC 217  
DB AC 87152  
218 ATTCAAGTAAAGTAACTACTATCTCAGAACTTA 249  
DB AATTAATGGCTACATTTTGTTTTAACTAA 87120

RESULT 6  
US-11-266-748A-23437

Sequence 23437, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 23437  
LENGTH: 162354  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-23437

Query Match 12.2%; Score 48.6; DB 8; Length 162354;  
Best Local Similarity 68.0%; Pred. No. 0.0036;  
Matches 66; Conservative 1; Mismatches 30; Indels 0; Gaps 0;

127 TTTTGTACTGTTGATATATATATATATATATATATATATATATATATATAT 186  
DB TGTGTGTGTGTAT 144910  
187 TTTTAATCTGGGACATAATATATATATATATATATATATATATATATATAT 223  
DB TTTTAATCTGGGACATAATATATATATATATATATATATATATATATATATAT 144970  
ACATTCTACCAACACACATATATATATATATATATATATATATATATATATATAT 145006

RESULT 7  
US-11-266-748A-223388

Sequence 223388, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:

APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 223388  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-223388

Query Match 12.1%; Score 48.4; DB 8; Length 1000;  
Best Local Similarity 86.7%; Pred. No. 0.0058;  
Matches 52; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

130 TTGTACTGTTGATATATATATATATATATATATATATATATATATATAT 189  
DB TGTGTGTGTGTAT 526  
TGTGTGTGTGTAT 585

RESULT 8

Sequence 290883, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18



NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 290883  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-290883

Query Match 12.1% Score 48.4; DB 8; Length 1000;  
Best Local Similarity 86.7%; Pred. No. 0.00058;  
Matches 52; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

130 TTGTTCTGTTGATATATATATATATACACACACACACACACACATATTTT 189  
DB 526 TTGTTCTCATATATATATATATATACACACACACACACACACATATTTT 585

RESULT 9  
US-11-266-748A-342312/c  
Sequence 342312, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 342312  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-342312

Query Match 12.1% Score 48.4; DB 8; Length 1000;  
Best Local Similarity 86.7%; Pred. No. 0.00058;  
Matches 52; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

130 TTGTTCTGTTGATATATATATATATACACACACACACACACACATATTTT 189  
DB 475 TTGTTCTCATATATATATATATATACACACACACACACACACATATTTT 416

RESULT 10  
US-11-266-748A-402247  
Sequence 402247, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same

FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 402247  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-402247

Query Match 12.1% Score 48.4; DB 8; Length 1000;  
Best Local Similarity 86.7%; Pred. No. 0.00058;  
Matches 52; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

130 TTGTTCTGTTGATATATATATATATACACACACACACACACACATATTTT 189  
DB 526 TTGTTCTCATATATATATATATATACACACACACACACACACATATTTT 585

RESULT 11  
US-11-266-748A-473293/c  
Sequence 473293, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
APPLICANT: Mulligan, Karl  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 473293  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-473293



	Query Match	12.1%	Score 48.4	DB 8	Length 1000;	
	Best Local Similarity	86.7%	Pred. No. 0.00058			
	Matches	52;	Conservative 1;	Mismatches 7;	Indels 0;	Gaps 0;
Oy	130	TTCAGCTTGATATATATATATATACACACACACACACACATATT				189
Dh	475	TTCGTTTCAACATATATATATATACACACACACACACACACATATT				416

RESULT 12  
US-11-266-748A-24706  
Sequence 24706, Application US/11266748A  
Publication NO. US2006013463A1  
GENERAL INFORMATION:  
APPLICANT: Hartin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcription Microarray Technology and  
TITLE OF INVENTION: Methods of Using the Same  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT APPLICATION NUMBER: US/11/266,748A  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 24706  
LENGTH: 2950  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-24706

```

Query Match Summary          12.0% Score 48; DB 8; Length 2950;
Best Local Similarity       73.2%; Pred. No. 0.0011;
Matches    60; Conservative   1; Mismatches    21; Indels     0; Gaps      0

QY      106 GCAGGGCAGTAGGCGCTACCTTTTGTACTGTTTGATATATATATATACACACAC 165
Db      2425 GCTGTCATATTAAAGATATGTGATATATATATATATATATATATATATATATACATATACAC 2485
              |||||
OY      166 ACACACACACACACACATATAT 187
Db      2485 ACACACACACACACACATATAT 2506
              |||||

RESULT 13
US-11-266-748A--203031/c
# Sequence 203031, Application US/11266748A
# Publication No. US2006013463A1
# GENERAL INFORMATION:
# APPLICANT: Harkin, Paul
# APPLICANT: Johnston, Patrick
# APPLICANT: Mulligan, Karl
# TITLE OF INVENTION: Transcription Microarray Technology and
# TITLE OF INVENTION: Methods of Using the Same
# FILE REFERENCE: 55815-0102 (319189)
# CURRENT APPLICATION NUMBER: US/11/266,748A
# CURRENT FILING DATE: 2005-11-03
# PRIOR APPLICATION NUMBER: EP 04105479.2

```

```

; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,223
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO: 203031
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-203031

```

Query Match	11.9%	Score 47.4	DB 8	Length 1000
Best Local Similarity	62.6%	Pred. No. 0.0011,		
Matches	72	Mismatches 42	Indels 0	Gaps 0
Oy	123	TACCTTTTGAGCTGGTATATATATATATACACACACACACACACA	182	
Dd	730	TATATTATATATATATATATATATATATATATATATACACACACACACACA	671	
Oy	183	YATATTTAACTCGGACATTAATGTACACCACCTATTCAAAAGTAAAACTACTA	237	
Dd	670	CACACACCTCATTTGAAAACATGAAGAAAACATTTAAAAATKCATACATCCTTTTA	616	

```

RESULT 14
US-11-266-748A-282791
; Sequence 282791, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Hartin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcription Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 282791
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-282791

```











APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 281212  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-281212

Query Match 11.7%; Score 46.6; DB 8; Length 1000;  
Best Local Similarity 70.1%; Pred. No. 0.0018;  
Matches 61; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 142 ATATATATATATATACACACACACACACACACATATTTTAACTGGGACA 201  
DB 866 ATATATATATATACACACACACACATATATATATATATATATATATACACACA 807

QY 202 TAAATGTACACACCTTATTCAGTAA 228  
DB 806 TACATACACACATGCTTTAAATAATCA 780

RESULT 21  
US-11-266-748A-307852  
Sequence 307852, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293

PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 307852  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-307852

Query Match 11.7%; Score 46.6; DB 8; Length 1000;  
Best Local Similarity 70.1%; Pred. No. 0.0018;  
Matches 61; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 142 ATATATATATATATACACACACACACACACACATATTTTAACTGGGACA 201  
DB 135 ATATATATATATACACACACACACATATATATATATATATATATATACACACA 194

QY 202 TAAATGTACACACCTTATTCAGTAA 228  
DB 195 TACATACACACATGCTTTAAATAATCA 221

RESULT 22  
US-11-266-748A-390020/c  
Sequence 390020, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 390020  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-390020

Query Match 11.7%; Score 46.6; DB 8; Length 1000;  
Best Local Similarity 70.1%; Pred. No. 0.0018;  
Matches 61; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 142 ATATATATATATATACACACACACACACACACATATTTTAACTGGGACA 201  
DB 866 ATATATATATATACACACACACACATATATATATATATATATATATACACACA 807

QY 202 TAAATGTACACACCTTATTCAGTAA 228  
DB 806 TACATACACACATGCTTTAAATAATCA 780

RESULT 23  
US-11-266-748A-390020  
Sequence 390020, Application US/11266748A  
Publication No. US20060134663A1  
GENERAL INFORMATION:  
APPLICANT: Harkin, Paul  
APPLICANT: Johnston, Patrick  
TITLE OF INVENTION: Transcriptome Microarray Technology and  
FILE REFERENCE: 55815-0102 (319189)  
CURRENT FILING DATE: 2005-11-03  
PRIOR APPLICATION NUMBER: US/11/266,748A  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105479.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105482.6  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105483.4  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105507.0  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105485.9  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: EP 04105484.2  
PRIOR FILING DATE: 2004-11-03  
PRIOR APPLICATION NUMBER: US 60/662,276  
PRIOR FILING DATE: 2005-03-14  
PRIOR APPLICATION NUMBER: US 60/700,293  
PRIOR FILING DATE: 2005-07-18  
NUMBER OF SEQ ID NOS: 483996  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 390020  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-266-748A-390020







[illegible]



```

? SEQ ID NO 525
? LENGTH: 12423
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURES:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
? HS-10-317-441-525

```

```
Query Match      11.5%;  Score 46;  DB 6;  Length 12423;
Best Local Similarity 59.4%;  Pred. No. 0.0071;
Matches 76;  Conservative 1;  Mismatches 51;  Indels 0;  Gaps 0;
```

QY	Db	QY	Db
139	TTGGTATATATATATACATACACACACACACACACATATTTTAAATCCGG	199	ACATAAATGTACACACACTATTTCAAAGTAAATCTCATCTCGAATTAAGCATAT
7720	TTACAAACCTTCT	7660	ACCTAAATCTAAATCTCTCTCTTAATAAACTTAATCTCTTAATAAAACCAAAAA
259	CTGTCTAT	259	CTGTCTAT
7600	CAAAATAT	7600	CAAAATAT

RESULT 29  
US-11-266-748A-23687/c  
Sequence 23687 Amplification ITS/11266748A

Query Match	11.5%	Score 46;	DB 8;	Length 118384;
Best Local Similarity	64.4%;	Pred. No. 0.017;		
Matches 67;	Conservative 1;	Mismatches 36;	Indels 0;	Gaps 0

Oy	153	TACATACACACACACACACACACAVATTTTTAACTCGGGACATAAATGTACAC	212
Db	115758	TACACACACACACACACACACACACCAATTTATCAAATGTATATACAGTCAT	115699
Oy	213	CACCATTTCCAAGTAAAACTACTATCTCAGAAGTAAAGACATA	256
Db	115698	GTTGTCATTCAATGCACAAGGATACCTTCTGAAGAAATGTATCCCTTA	115655

```

RESULT 30
US-10-533-365-1/c
; Sequence 1, Application US/10533365
; Publication NO. US2006014162A1
; GENERAL INFORMATION:
; APPLICANT: Reynoldsdottlr, Inga
; APPLICANT: Gulcher, Jeffrey R.
; APPLICANT: Grant, Struan F.
; APPLICANT: Thorleifsson, Gudmar
; TITLE OF INVENTION: Human Type II Diabetes Gene - Sltt-3
; TITLE OF INVENTION: Located on Chromosome 5q35
; FILE REFERENCE: 2345.2046-007
; CURRENT APPLICATION NUMBER: US/10/533,365
; CURRENT FILING DATE: 2005-04-29
; PRIOR APPLICATION NUMBER: PCT/US03/34801
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 60/423,541
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 236
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 634888
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-533-365-1

```

```

Query Match      11.5%; Score 46; DB 6; Length 634880;
Best Local Similarity 62.5%; Pred. No. 0.032;
Matches 70; Conservative 1; Mismatches 41; Indels 0; Gaps 0;

```

QY 95 AAAGTAGATATCAGGGCAGAGGAGCCATACCTTTTGTACGTTTGAAATATATATA 154

Db 102116 AAACATATACACACACACACACATATGTGTATTTTATATATATATATA 102057

QY 155 CATACACACACACACACACACATATTTTAACTCGGCATATAAT 206

Db 102056 AATACACACACACACACACACATATACATATACACATGCTCTCGGT 102005

```

RESULT 31
US-11-266-748A-293635/C
/ Sequence 293635, Application US/11266748A
/ Publication No. US2006013463A1
/ GENERAL INFORMATION:
/ APPLICANT: Harkin, Paul
/ APPLICANT: Johnston, Patrick
/ APPLICANT: Mulligan, Karl
/ TITLE OF INVENTION: Transcription Microarray Technology and
/ TITLE OF INVENTION: Methods of Using the Same
/ FILE REFERENCE: 55815-0102 (319189)
/ CURRENT APPLICATION NUMBER: US/11/266,748A
/ CURRENT FILING DATE: 2005-11-03
/ PRIOR APPLICATION NUMBER: EP 04105479.2
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105482.6
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105483.4
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105507.0
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105485.9
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: EP 04105484.2
/ PRIOR FILING DATE: 2004-11-03
/ PRIOR APPLICATION NUMBER: US 60/662,276
/ PRIOR FILING DATE: 2005-03-14
/ PRIOR APPLICATION NUMBER: US 60/700,293
/ PRIOR FILING DATE: 2005-07-18
/ NUMBER OF SEQ ID NOS: 48396
/ SOFTWARE: Patclicin version 3.3
/ SEQ ID NO 293635

```















```

; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 48396
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 287071
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-287071

Query Match      11.4%; Score 45.6; DB 8; Length 1000;
Best Local Similarity 77.1%; Pred. No. 0.0035;
Matches 54; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 136 TGGTTGATATATATATACACACACACACACACACATATTTTAACTC 195
Db 299 TGAATTTAAAAAATATATATATACACACACACACACACATTTTATC 240

QY 196 GGGACATMAA 205
Db 239 AGGAGAGAAA 230

RESULT 41
US-11-266-748A-290324
; Sequence 290324, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcription Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 48396
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 290324
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-290324

Query Match      11.4%; Score 45.6; DB 8; Length 1000;
Best Local Similarity 88.9%; Pred. No. 0.0035;
Matches 48; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 136 TGGTTGATATATATATACACACACACACACACACATATTTT 169
Db 577 TAAATTAATATATATATATATACACACACACACACACATATAT 630

RESULT 42
US-11-266-748A-338500
; Sequence 338500, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcription Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 48396
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 338500
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-266-748A-338500

Query Match      11.4%; Score 45.6; DB 8; Length 1000;
Best Local Similarity 77.1%; Pred. No. 0.0035;
Matches 54; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 136 TGGTTGATATATATATATACACACACACACACACACATATTTTAACTC 195
Db 702 TGAATTTAAAAAATATATATATATACACACACACACACATTTTATC 761

QY 196 GGGACATMAA 205
Db 762 AGGAGAGAAA 771

RESULT 43
US-11-266-748A-341753/C
; Sequence 341753, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcription Microarray Technology and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; CURRENT FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
```



```
;; PRIOR FILING DATE: 2005-07-18
;; NUMBER OF SEQ ID NOS: 483996
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 341753
;; LENGTH: 1000
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
US-11-266-748A-341753

Query Match
Best Local Similarity 11.4%; Score 45.6; DB 8; Length 1000;
Matches 48; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 136 TGTTCATATATATATACACACACACACACACACATATTT 189
DB 424 TATTTATATATATATATATATACACACACACACACACACATATAT 371

RESULT 44
US-11-266-748A-397664/C
;; Sequence 397664, Application US/11266748A
;; Publication No. US2006013463A1
;; GENERAL INFORMATION:
;; APPLICANT: Hartkin, Paul
;; APPLICANT: Johnston, Patrick
;; APPLICANT: Mulligan, Karl
;; TITLE OF INVENTION: Transcriptome Microarray Technology and
;; TITLE OF INVENTION: Methods of Using the Same
;; FILE REFERENCE: 55815-0102 (319189)
;; CURRENT APPLICATION NUMBER: US/11/266,748A
;; PRIOR FILING DATE: 2005-11-03
;; PRIOR APPLICATION NUMBER: EP 04105479.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105482.6
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105483.4
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105507.0
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105485.9
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105484.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 60/662,276
;; PRIOR FILING DATE: 2005-03-14
;; PRIOR APPLICATION NUMBER: US 60/700,293
;; PRIOR FILING DATE: 2005-07-18
;; NUMBER OF SEQ ID NOS: 483996
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 397664
;; LENGTH: 1000
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
US-11-266-748A-397664

Query Match
Best Local Similarity 11.4%; Score 45.6; DB 8; Length 1000;
Matches 54; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
```

```
;; APPLICANT: Hartkin, Paul
;; APPLICANT: Johnston, Patrick
;; APPLICANT: Mulligan, Karl
;; TITLE OF INVENTION: Transcriptome Microarray Technology and
;; TITLE OF INVENTION: Methods of Using the Same
;; FILE REFERENCE: 55815-0102 (319189)
;; CURRENT APPLICATION NUMBER: US/11/266,748A
;; PRIOR FILING DATE: 2005-11-03
;; PRIOR APPLICATION NUMBER: EP 04105479.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105482.6
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105483.4
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105507.0
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105485.9
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105484.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: US 60/662,276
;; PRIOR FILING DATE: 2005-03-14
;; PRIOR APPLICATION NUMBER: US 60/700,293
;; PRIOR FILING DATE: 2005-07-18
;; NUMBER OF SEQ ID NOS: 483996
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 468710

RESULT 46
US-11-266-748A-468710
;; Sequence 468710, Application US/11266748A
;; Publication No. US2006013463A1
;; GENERAL INFORMATION:
;; APPLICANT: Hartkin, Paul
;; APPLICANT: Johnston, Patrick
;; APPLICANT: Mulligan, Karl
;; TITLE OF INVENTION: Transcriptome Microarray Technology and
;; TITLE OF INVENTION: Methods of Using the Same
;; FILE REFERENCE: 55815-0102 (319189)
;; CURRENT APPLICATION NUMBER: US/11/266,748A
;; PRIOR FILING DATE: 2005-11-03
;; PRIOR APPLICATION NUMBER: EP 04105479.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105482.6
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105483.4
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105507.0
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105485.9
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: EP 04105484.2
;; PRIOR FILING DATE: 2004-11-03
;; PRIOR APPLICATION NUMBER: US 60/662,276
;; PRIOR FILING DATE: 2005-03-14
;; PRIOR APPLICATION NUMBER: US 60/700,293
;; PRIOR FILING DATE: 2005-07-18
;; NUMBER OF SEQ ID NOS: 483996
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 468710

QY 136 TGTTCATATATATATATACACACACACACACACACATATTT 189
DB 577 TATTTATATATATATATATATACACACACACACACACACATATAT 630
```











